International Journal of Approximate Reasoning 136 (2021) 249–280

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Information graphs and their use for Bayesian network graph construction

Remi Wieten [a,*], Floris Bex [a,b], Henry Prakken [a,c], Silja Renooij [a]

[a] Department of Information and Computing Sciences, Utrecht University, the Netherlands
[b] Institute for Law, Technology and Society, Tilburg University, the Netherlands
[c] Faculty of Law, University of Groningen, the Netherlands

ARTICLE INFO

Article history:
Received 5 June 2020
Received in revised form 13 April 2021
Accepted 17 June 2021
Available online 25 June 2021

Keywords:
Bayesian networks
Causal and evidential reasoning
Deduction
Abduction
Uncertainty
Qualitative probabilistic reasoning

ABSTRACT

In this paper, we present the information graph (IG) formalism, which provides a precise account of the interplay between deductive and abductive inference and causal and evidential information, where 'deduction' is used for defeasible 'forward' inference. IGs formalise analyses performed by domain experts in the informal reasoning tools they are familiar with, such as mind maps used in crime analysis. Based on principles for reasoning with causal and evidential information given the evidence, we impose constraints on the inferences that may be performed with IGs. Our IG-formalism is intended to facilitate the construction of formal representations within AI systems by serving as an intermediary formalism between analyses performed using informal reasoning tools and formalisms that allow for formal evaluation. In this paper, we investigate the use of the IG-formalism as an intermediary formalism in facilitating Bayesian network (BN) graph construction. We propose a structured approach for automatically constructing from an IG a directed BN graph, together with qualitative constraints on the probability distribution represented by the BN. Moreover, we prove a number of formal properties of our approach and identify assumptions under which the construction of an initial BN graph can be fully automated.

1. Introduction

Bayesian networks (BNs) [20] are compact graphical representations of joint probability distributions that have found applications in many fields where uncertainty and evidence plays a role, including medicine, engineering, forensics and law [16]. For instance, in recent years legal and forensic experts have increasingly developed and used BNs for the interpretation of different types of forensic trace evidence [36], such as glass fragments, DNA traces and finger marks [34], as well as for modelling crime linkage [44]. A BN consists of a directed acyclic graph (DAG), which captures the probabilistic independence relation among variables relevant to the domain, and locally specified (conditional) probability distributions that collectively describe a joint probability distribution. BNs are well-suited for reasoning about the uncertain consequences that can be inferred from evidence. Domain experts, however, typically do not have the expertise to construct mathematical models and misinterpret the directed arcs of a BN as non-symmetric relations of cause and effect instead of collectively encoding an independence relation [12]. Especially in data-poor domains, BN construction therefore needs to be done mostly manually through a knowledge elicitation procedure in consultation with the domain expert, which is a difficult and error-prone task

* Corresponding author.
E-mail address: g.m.wieten@uu.nl (R. Wieten).

https://doi.org/10.1016/j.ijar.2021.06.007
0888-613X/

[18], and domain experts resort to using other reasoning tools including mind maps [26,38], argument diagrams [1,3,26], Wigmore charts [43], and ontologies [17,31], as well as basic text-based tools such as Microsoft Word and Excel [26].

Methods have been proposed that facilitate BN construction by extracting information relevant for a BN from analyses performed in tools used by experts. For instance, methods for constructing BNs from information represented in ontologies, knowledge representations which capture relations between concepts in a domain, have been proposed [17,31]. To apply these approaches in practice, the problem under consideration first needs to be specified in the formal ontology language required as input. Aforementioned tools such as mind maps, argument diagrams and Wigmore charts similarly do not directly allow for guiding BN construction due to their informal nature. In contrast with ontologies, these tools are used to capture inferences made with causal and evidential information (see [2,6]), instead of with generic relations between concepts. In this paper, we focus on reasoning tools such as mind maps, where we wish to formalise analyses performed using such tools in a manner that (1) adheres to principles from the literature on reasoning with causal and evidential information [2,21,27,29], while (2) allowing inference to be performed in a manner closely related to the way in which inference is performed using such tools, and that (3) allows for guiding BN construction.

Principles from the literature on reasoning with causal and evidential information state that inference is often performed using domain-specific *generalisations* [1,2], also called defaults [27,32], which capture knowledge about the world in conditional form. We distinguish between *causal* generalisations (e.g. fire typically causes smoke) and *evidential* generalisations (e.g. smoke is evidence for fire) [2,27]. Inference can be performed in a *deductive* or forward fashion, where from a generalisation (e.g. fire typically causes smoke) and its antecedent (fire), the consequent (smoke) is inferred, and in an *abductive* [21] or backward fashion, where from a causal generalisation and by affirming the consequent the antecedent is inferred. Note that the term 'deduction' is not consistently used in the literature, as it can either mean strict inference, in which the consequent universally holds given the antecedents (e.g. [30]), or defeasible inference, in which the consequent tentatively holds given the antecedents (e.g. [35]). In this paper, 'deduction' is used for defeasible 'forward' inference.

When performing analyses in aforementioned reasoning tools such as mind maps, domain experts naturally mix both causal and evidential generalisations and perform both deductive and abductive inferences, where the used generalisations and the inference type (deduction, abduction) are typically left implicit. Hence, in formalising analyses performed using these tools we need a precise account of the interplay between the different types of inferences and generalisations and the constraints on performing inference we need to impose. In this paper we present the information graph (IG) formalism [42], which provides such an account. IGs are knowledge representations that formalise analyses performed by domain experts using the informal reasoning tools they are familiar with in a manner that makes the causal and evidential generalisations used in performing inference explicit. Based on principles for reasoning with causal and evidential generalisations, we then define how deductive and abductive inference can be performed with IGs given a set of propositions labelled evidence. Most existing formalisms that allow both inference types with causal and evidential information are logic-based (e.g. [2,29,35]); instead, we prefer a graph-based formalism to remain closely related to analyses performed using aforementioned graph-based tools as well as the BN-formalism.

Our IG-formalism is intended to facilitate the construction of formal representations within AI systems by serving as an intermediary formalism between analyses performed using informal reasoning tools and formalisms that allow for formal evaluation. In this paper, we investigate the use of the IG-formalism as an intermediary formalism in facilitating BN graph construction. We propose a structured approach for automatically constructing a directed BN graph from an IG by exploiting the causal and evidential knowledge expressed in an IG. In manual BN graph construction, the notion of causality is commonly used as a guiding principle [16,20] instead of directly eliciting conditional independencies. In IGs, causality information is made explicit and can thus be directly used in BN graph construction. In addition, we demonstrate that the inferences that can be read from an IG given the evidence provide for qualitative constraints on the probability distribution represented by the BN. We formally prove that BN graphs constructed by our approach capture reasoning patterns similar to those represented by the original IG. Moreover, we identify assumptions under which the fully automatically constructed initial graph is guaranteed to be a DAG, and identify bounds on the complexity of probabilistic inference in BNs constructed by our approach.

The IG-formalism as presented in this paper is a further specification of the IG-formalism that appeared in our previous work [42], in which the relations between inference as it can be performed with IGs and argumentation were investigated. Argumentation [15,30] is particularly suited for adversarial settings such as the legal domain, where arguments for and against claims are constructed from evidence. In [42], it is shown that an Argumentation Framework (AF) as in Dung [15] can straightforwardly be generated from an IG by considering the available evidence, which allows arguments to be formally evaluated. The BN graph construction approach as presented in the current paper extends on our previous work on facilitating BN graph construction [5,39,40]; further details of this work are discussed in Sect. 9.

The paper is structured as follows. In Sect. 2 we provide principles for reasoning with causal and evidential information. In Sect. 3 we present an example of an analysis performed using a tool typically used by domain experts, namely a mind mapping tool, which illustrates that both deduction and abduction are performed by domain experts, using both causal and evidential generalisations. Based on this example, in Sect. 4 we motivate and define our IG-formalism. Sections 5 to 8 concern the construction of BNs from IGs. Section 5 provides preliminaries on BNs. In Sect. 6 we present our approach for constructing BN graphs from IGs. In Sect. 7 we prove formal properties of our approach. In Sect. 8 we illustrate and perform a first validation of our approach by applying it to parts of an actual legal case, namely the well-known Sacco and Vanzetti case, where we compare our results to a previous BN modelling of the same case [22]. In Sect. 9 related research on among others inference with causality information and BN graph construction is discussed. In Sect. 10 we discuss future work, summarise our findings, and conclude.

2. Reasoning with causal and evidential information

In this section, we provide principles for reasoning with causal and evidential information, where we review the terminology used to describe it and introduce assumptions that demarcate the scope of the work presented in this paper. Inference is the process of drawing conclusions from premises starting from the evidence, where evidence is that what has been established with certainty in the context under consideration. For instance, in the context of a legal trial the evidence consists of that what is actually observed by a judge or jury, such as documents (e.g. police and autopsy reports) and other tangible evidence, as well as testimonial evidence [2]. Inference is often performed using domain-specific *generalisations* [1,2], also called defaults [27,32], which capture knowledge about the world in conditional form. We distinguish between causal and evidential generalisations [2,27]. Causal generalisations are of the form '$c_1, \dots, c_n$ *usually/normally/typically cause e*' (e.g. 'fire typically causes smoke') and evidential generalisations are of the form '$e_1, \dots, e_n$ *are evidence for c*' (e.g. 'smoke is evidence for fire'). We denote generalisations as *fire* $\rightarrow$ *smoke*, where *fire* is the generalisation's *antecedent* and *smoke* its *consequent*. For a causal generalisation, its antecedents express a cause for the consequent, and for an evidential generalisation, its consequent expresses the usual cause for its antecedents. We assume that generalisations have one or more antecedents and exactly one consequent. In case a generalisation has multiple antecedents, it expresses that only the antecedents together allow us to infer the consequent. The notation $\rightarrow_C$ and $\rightarrow_E$ is used for causal and evidential generalisations, respectively.

Different types of inferences can be performed with generalisations depending on whether their antecedents or consequent are *affirmed* in that they are either observed or inferred; here, a consequent or antecedent is *inferred* iff it is either deductively or abductively inferred.

2.1. Deductive inference

Inference can be performed in a *deductive* fashion, where from a causal or evidential generalisation and by affirming the antecedents, the consequent is inferred by modus ponens on the generalisation. As noted in the introduction, the term 'deduction' is used for defeasible 'forward' inference; hence, deduction is not a stronger or more reliable form of inference than abduction, which is another type of defeasible inference. *Prediction* [35] is a specific type of deductive inference in which the consequent of a causal generalisation is deductively inferred by affirming its antecedents.

2.2. Abductive inference

Abduction [21] can also be performed: from a causal generalisation and by affirming the consequent, the antecedents are inferred, since if the antecedents are true it would allow us to deductively infer the consequent modus-ponens-style. In case multiple causes for a common effect are abductively inferred using multiple causal generalisations with the same consequent, then these causes are considered to be *competing alternative explanations* [21] for the common effect expressed by the consequent. In case a causal generalisation has multiple antecedents, we assume that these antecedents are not in competition among themselves.

Example 1. Consider the causal generalisations *fire* $\rightarrow_C$ *smoke* and *smoke_machine* $\rightarrow_C$ *smoke*. By affirming the common consequent (*smoke*), *fire* and *smoke_machine* are abductively inferred, which are then competing alternative explanations of *smoke*. □

2.3. Representing causal knowledge

Abductive inference with causal generalisations and deductive inference with evidential generalisations are related: in some cases, we will accept not only causal generalisation '*c usually/normally/typically causes e*' but also evidential generalisation '*e is evidence for c*' [4,27], which we will call the *evidential counterpart* of the causal generalisation. However, it can be argued that we only accept the evidential counterpart of a causal generalisation if c is the *usual* cause of e, where we assume that only one cause can be the usual cause of e.

Example 2. Fire can be considered the usual cause of smoke, so we will accept both causal generalisation g: *fire* $\rightarrow_C$ *smoke* and its evidential counterpart g': *smoke* $\rightarrow_E$ *fire*. In this case, abductive inference with generalisation g can be encoded as deductive inference with generalisation g'. Because a smoke machine cannot be considered the usual cause of smoke, we will accept causal generalisation *smoke_machine* $\rightarrow_C$ *smoke* but we will not accept evidential generalisation *smoke* $\rightarrow_E$ *smoke_machine*. □

Note that a causal generalisation g can only have an evidential counterpart g' in case g has a single antecedent, as we assume generalisations have a single consequent but multiple antecedents. Furthermore, as we assume that only one cause can be the usual cause of e, only one of the causal generalisations $c_1 \rightarrow_c e$ or $c_2 \rightarrow_c e$ can be replaced by an evidential generalisation. Hence, we do not consider c_1 and c_2 to be competing alternative explanations of e in case deduction is performed using evidential generalisations $e \rightarrow_e c_1$ and $e \rightarrow_e c_2$.

2.4. Mixed inference

Deduction and abduction can be iteratively performed, where *mixed* abductive-deductive inference is also possible.

Example 3. Suppose that from the causal generalisation *fire* $\rightarrow_C$ *smoke* and by affirming its consequent (*smoke*), its antecedent (*fire*) is inferred. Now, if the additional causal generalisation *fire* $\rightarrow_C$ *heat* is provided, then its consequent (*heat*) can be deductively inferred (or predicted) as the antecedent (*fire*) has been previously abductively inferred. □

Mixed deductive inference, using both causal and evidential generalisations, can also be performed [4], but as noted by Pearl [27] care should be taken in performing mixed inference that no cause for an effect is inferred in case an alternative cause for this effect was already previously inferred.

Example 4. (a) Consider the example in which a causal generalisation *smoke_machine* $\rightarrow_C$ *smoke* and an evidential generalisation *smoke* $\rightarrow_E$ *fire* are provided. Deductively chaining these generalisations would make us infer that there is a fire when seeing a smoke machine, which is clearly undesirable.

(b) Similarly, in performing mixed deductive-abductive inference, care should be taken that no cause for an effect is inferred in case an alternative cause for this effect was already previously inferred. Consider the above example, where instead of an evidential generalisation *smoke* $\rightarrow_E$ *fire* a causal generalisation *fire* $\rightarrow_C$ *smoke* is now provided. Upon seeing a smoke machine, this would make us infer that there is a fire in case deductive inference and abductive inference are performed in sequence, which is again undesirable. □

Accordingly, we wish to prohibit these types of inference patterns, and refer to the constraint that no cause for an effect should be inferred in case an alternative cause for this effect was already previously inferred as *Pearl's constraint* [27].

2.5. Ambiguous inference

Finally, situations may arise in practice in which both deductive and abductive inference can be performed with the same causal generalisation; the inference type is, therefore, considered *ambiguous*.

Example 5. Consider the causal generalisation *fire* $\rightarrow_C$ *smoke* and assume that both *fire* and *smoke* are affirmed but not observed, then both deductive and abductive inference can be performed to either infer *smoke* from *fire* or *fire* from *smoke*, respectively. □

3. Example of an analysis performed using a mind mapping tool

In this section, we present an example of an analysis performed using a mind mapping tool [26], which is an example of a tool typically used by domain experts, for instance in crime analysis [38]. Based on this example, we motivate and illustrate the design choices for our IG-formalism in Sect. 4. A mind map usually takes the shape of a diagram in which hypotheses and claims are represented by boxes and underlined text, and undirected edges symbolise relations between these hypotheses and claims. An example is depicted in Fig. 1, which is based on a standard template used by the Dutch police for criminal cases involving the suspicious death of a person. The mind map represents various scenario-elements and the crime analyst uses evidence to support or oppose these elements, indicated in the mind map by plus and minus symbols, respectively.

Example 6. An example of a partially filled in mind map is depicted in Fig. 1, which also serves as our running example. In this example case, adapted from [2], the high-level hypothesis 'Murder' is considered; for illustration purposes the details of the case have been changed. The case concerns the murder of Leo de Jager, which took place in the small Dutch town of Anjum. Leo's body was found on the property of Marjan van der E.; we are interested in her involvement in the murder. As a police report (*police_report*) indicates that Leo's body was found on Marjan's property, the claim *marjan_murdered_leo* is added as an answer to the 'Who' question. By means of a plus symbol and an undirected edge connecting the evidence to the claim, it is indicated that the police report supports the claim that Marjan murdered Leo. Possible motives (*motive_1* and *motive_2*) are provided as to why Marjan may have wanted to murder Leo, which are connected to the 'Why' question via undirected edges. Claims *testimony_1* and *testimony_2* support these two motives, indicated by the plus symbols connected to these claims. In her testimony (*testimony_3*), Marjan denied any involvement in the murder of Leo, which is indicated by a minus symbol. This opposes the claim that Marjan murdered Leo. Further testimony (*testimony_4*) indicates that Marjan had reason to lie when giving her testimony (*lie*). By means of a minus symbol and an undirected edge connecting *lie* to

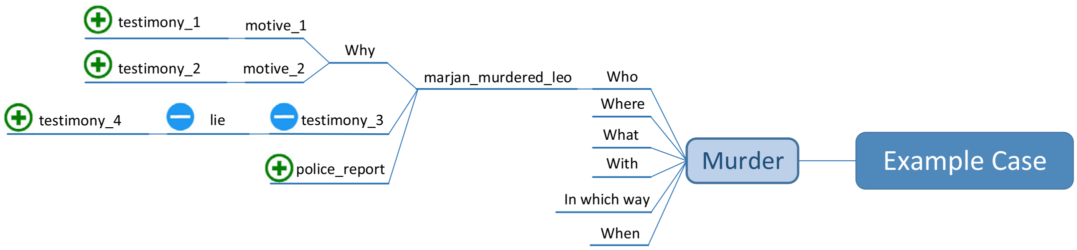


Fig. 1. Example of a partially filled in mind map.

testimony_3, it is indicated that this claim weakens the inference step from her testimony to the claim that she did not murder Leo. □

As the edges in a mind map are undirected, it is unclear from this graphical representation alone which types of generalisations and inferences were used in constructing this map. Establishing this with certainty would require directly consulting the domain experts involved in constructing the chart. We note, however, that the reasoning performed in constructing this mind map can be interpreted in multiple ways. One interpretation is that the domain expert first (preliminarily) inferred that Marjan murdered Leo from the police report via deductive inference using the evidential generalisation *police_report* $\rightarrow_{\mathrm{e}}$ *marjan_murdered_leo*, and then abductively inferred the two possible motives using the causal generalisations g_i: *motive_i* $\rightarrow_{\mathrm{c}}$ *marjan_murdered_leo*; $i = 1, 2$. These two causes are then competing alternative explanations as to why Marjan murdered Leo and are subsequently grounded in evidence, namely via deductive inference from the testimonies using evidential generalisations g'_j: *testimony_j* $\rightarrow_{\mathrm{e}}$ *motive_j*; $j = 1, 2$. An alternative interpretation is that the mind map was constructed iteratively from the evidence, where from the testimonies the motives are inferred via deductive inference using generalisations g'_1 and g'_2. The claim that Marjan murdered Leo is then inferred modus-ponens style: from causal generalisations g_1 and g_2 and the previously inferred antecedents, the consequent is deductively inferred. In this way, the two motives are not in competition for the common effect that Marjan murdered Leo.

The above example illustrates that the types of generalisations and inferences that are involved in the analysis of a case using a mind mapping tool are typically left unspecified. Similarly, in mind maps the exact manner in which claims and links conflict is not precisely specified: a minus symbol can either indicate support for the opposing claim (e.g. *testimony_3* supports the negation of *marjan_murdered_leo*) or indicate an exception to the performed inference (e.g. *lie* opposes the inference from *testimony_3* to the negation of *marjan_murdered_leo*).

4. The information graph formalism

The example from Sect. 3 makes it plausible that both deduction and abduction are performed by domain experts when performing analyses using reasoning tools they are familiar with. In performing such analyses, the used generalisations, as well as the inference type (deduction, abduction), are left implicit. Furthermore, the assumptions of domain experts underlying their analyses are typically not explicitly stated, making these analyses ambiguous to interpret. For current purposes, we wish to provide a precise account of the interplay between the different types of inferences and generalisations that formalises and disambiguates these analyses in a manner that makes the used generalisations explicit. *Information graphs* (IGs) [42], which we define in Sect. 4.1, are knowledge representations that explicitly describe causal and evidential generalisations in the graph. In Sect. 4.2, we define how deductive and abductive inferences can be read from IGs given the evidence, based on the principles for reasoning with causal and evidential information discussed in Sect. 2.

4.1. Information graphs

IGs are defined as follows.

Definition 1 *(Information graph).* An *information graph* (IG) is a directed graph $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$, where $\mathbf{P}$ is a set of nodes representing propositions from a propositional literal language with ordinary negation symbol $\neg$. $\mathbf{A}_{\mathcal{I}}$ is a set of (hyper)arcs that divides into three pairwise disjoint subsets $\mathbf{G}$, $\mathbf{N}$ and $\mathbf{Exc}$ of generalisation arcs, negation arcs and exception arcs, defined in Definitions 2, 6, and 7, respectively.

We write $p = -q$ in case $p = \neg q$ or $q = \neg p$. Note that an IG $G_{\mathcal{I}}$ does not have to be a connected graph.

Definition 2 *(Generalisation arc).* Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. A *generalisation arc* $g \in \mathbf{G} \subseteq \mathbf{A}_{\mathcal{I}}$ is a directed (hyper)arc g: $\{p_1, \ldots, p_n\} \rightarrow p$, indicating a generalisation with antecedents $\mathbf{P_1} = \{p_1, \ldots, p_n\} \subseteq \mathbf{P}$ and consequent $p \in \mathbf{P} \setminus \mathbf{P_1}$. Here, propositions in $\mathbf{P_1}$ are called the *tails* of g, denoted by $\mathbf{Tails}(g)$, and p is called the *head* of g, denoted by $Head(g)$. $\mathbf{G}$ divides into two disjoint subsets $\mathbf{G}^c$ and $\mathbf{G}^e$ of causal and evidential generalisation arcs, respectively.

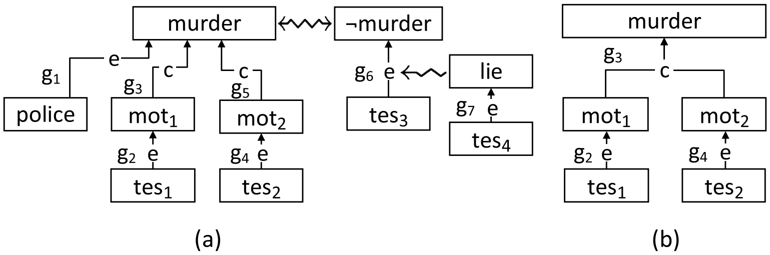


Fig. 2. An IG corresponding to a possible interpretation of the mind map of Fig. 1 (a); adjustment to the IG of Fig. 2a including generalisation arc $g_3 : \{mot_1, mot_2\} \to murder$ (b).

Curly brackets are omitted in case $|\mathbf{Tails}(g)| = 1$. In figures in this paper, generalisation arcs are denoted by solid (hyper)arcs, which are labelled 'c' for $g \in \mathbf{G}^c$ and 'e' for $g \in \mathbf{G}^e$.

A causal generalisation $g : c \to e$ may have an evidential counterpart of the form $g' : e \to c$ (see Sect. 2.3), but only if c is the usual cause of e. Definition 2 does not prohibit the coexistence of a causal generalisation $g : c \to e$ and its evidential counterpart $g' : e \to c$ in an IG, and inferences can be read from IGs including both generalisations without yielding anomalous results; hence, both generalisations may be included if this is considered desirable. However, it should be noted that g and g' represent the same knowledge, and that care should be taken in for instance modelling exceptions to generalisations (see Definition 7), as an exception to g can also be considered an exception to g'. Ultimately, it is the responsibility of the knowledge engineer in consultation with the domain expert to decide which knowledge to include in the IG and to ensure this knowledge is correctly and consistently represented.

In the following example, the mind map of Sect. 3 is modelled as an IG.

Example 7. In Fig. 2a, an IG is depicted for a possible interpretation of the running example. First, we consider the undirected edges connected to the testimonies and the police report in the mind map of Fig. 1. In an empirical study in the legal domain, van den Braak and colleagues [6] found that subjects often considered testimonies to be evidential, where generalisations are of the form *'Testimony to fact x is evidence for x'*. Police reports can similarly be considered evidential. The IG therefore includes generalisation arcs $g_1, g_2, g_4, g_7 \in \mathbf{G}^e$ to denote these generalisations. As tes_3 concerns Marjan's testimony to denying any involvement in the murder, $\neg murder$ is included in $\mathbf{P}$ and $g_6 :\ tes_3 \to \neg murder$ in $\mathbf{G}^e$. A motive for committing an act can be considered a cause for committing that act [6]. The IG therefore includes generalisation arcs $g_3 :\ mot_1 \to murder$ and $g_5 :\ mot_2 \to murder$ in $\mathbf{G}^c$ to denote these generalisations. □

Specific configurations of generalisation arcs express that two propositions are *alternative causes* of a common effect, as captured by the following definition.

Definition 3 *(Alternative causes).* Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. Then $c_1 \in \mathbf{P}$ and $c_2 \in \mathbf{P}$ are *alternative causes* of $e \in \mathbf{P}$, as indicated by generalisations g and g' in $\mathbf{G}$, iff one of the following holds:

1. $g \in \mathbf{G}^e$, $Head(g) = c_1$, $e \in \mathbf{Tails}(g)$, and either:
   1a) $g' \in \mathbf{G}^e$, $g' \neq g$, $Head(g') = c_2$, $e \in \mathbf{Tails}(g')$, or;
   1b) $g' \in \mathbf{G}^c$, $Head(g') = e$, $c_2 \in \mathbf{Tails}(g')$.

2. $g \in \mathbf{G}^c$, $Head(g) = e$, $c_1 \in \mathbf{Tails}(g)$, and either:
   2a) $g' \in \mathbf{G}^c$, $g' \neq g$, $Head(g') = e$, $c_2 \in \mathbf{Tails}(g')$, or;
   2b) $g' \in \mathbf{G}^e$, $Head(g') = c_2$, $e \in \mathbf{Tails}(g')$.

Generalisation chains are sequences of generalisation arcs.

Definition 4 *(Generalisation chain).* Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. Generalisation arcs $g_1, \ldots, g_m \in \mathbf{G} \subseteq \mathbf{A}_{\mathcal{I}}$ form a *generalisation chain* $[g_1, \ldots, g_m]$ in $G_{\mathcal{I}}$ iff $Head(g_{i-1}) \in \mathbf{Tails}(g_i)$ for $1 < i \leq m$.

Note that a subchain of a generalisation chain is again a generalisation chain.

Example 8. In the IG of Fig. 2a, $[g_2, g_3]$ is a generalisation chain as $Head(g_2) = mot_1 \in \mathbf{Tails}(g_3)$.

Consider the IG of Fig. 2b, which is an adjustment to the IG of Fig. 2a in which generalisation arc $g_3 : \{mot_1, mot_2\} \to murder$ in $\mathbf{G}^c$ is included instead of two separate generalisation arcs g_3 and g_5. According to Definition 4, $[g_2, g_3]$ is a

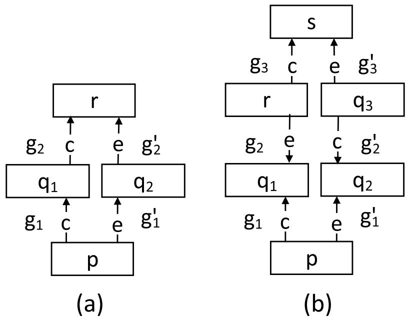


Fig. 3. Examples of IGs including causal cycles.

generalisation chain, but mot_2 is neither a head nor a tail of generalisation arc g_2; it suffices that $Head(g_2) = mot_1 \in$ $\mathbf{Tails}(g_3)$. $\square$

We define the following notion of a *causal cycle*.

Definition 5 *(Causal cycle).* Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. Proposition $p \in \mathbf{P}$ expresses a *direct cause* for $q \in \mathbf{P}$ iff $\exists g \in \mathbf{G} \subseteq \mathbf{A}_{\mathcal{I}}$ with $g \in \mathbf{G}^c$, $p \in \mathbf{Tails}(g)$, $q = Head(g)$ or $g \in \mathbf{G}^e$, $p = Head(g)$, $q \in \mathbf{Tails}(g)$. Proposition $p_1 \in \mathbf{P}$ expresses an *indirect cause* for $p_3 \in \mathbf{P}$ iff $\exists p_2 \in \mathbf{P}$, $p_2 \neq p_1$, $p_2 \neq p_3$, such that p_1 expresses a direct cause for p_2 and p_2 expresses a direct or indirect cause for p_3. A *causal cycle* exists in $G_{\mathcal{I}}$ iff $\exists p, q \in \mathbf{P}$ such that p expresses a direct or indirect cause for $q \in \mathbf{P}$ and q or $-q$ expresses a direct or indirect cause for p or for $-p$.

Examples of IGs including causal cycles are provided in Fig. 3.

We assume that graphs constructed in our IG-formalism conform to the following restrictions on generalisation chains, which arguably are reasonable rational constraints to impose on IGs. Informally, assumptions 1 and 2 exclude the possibility of using a proposition p to deductively infer itself or $-p$.

1. IGs only contain *non-repetitive* generalisation chains $[g_1, \ldots, g_m]$ in that $Head(g_m) \notin \mathbf{Tails}(g_1)$.
2. IGs only contain *consistent* generalisation chains $[g_1, \ldots, g_m]$ in that $\nexists i, j \in \{1, \ldots, m\}$ such that $Head(g_i) = -Head(g_j)$.
3. IGs do not include causal cycles (see also [29]).

A negation arc captures a conflict between a proposition and its negation[1] expressed in an IG.

Definition 6 *(Negation arc).* Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. A *negation arc* $n \in \mathbf{N} \subseteq \mathbf{A}_{\mathcal{I}}$ is a bidirectional arc $n\colon p \leftrightsquigarrow q$ in $G_{\mathcal{I}}$ that exists between a pair $p, q \in \mathbf{P}$ iff $q = -p$.

Example 9. Consider the running example. As both *murder* and $\neg$*murder* are included in the IG of Fig. 2a, negation arc n: *murder* $\leftrightsquigarrow$ $\neg$*murder* is also included in the graph. $\square$

As generalisations hardly ever hold universally, exceptional circumstances can be provided under which a generalisation may not hold; hence, we allow exceptions to generalisations to be explicitly specified in IGs.

Definition 7 *(Exception arc).* Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. An *exception arc* $exc \in \mathbf{Exc} \subseteq \mathbf{A}_{\mathcal{I}}$ is a hyperarc $exc\colon p \rightsquigarrow g$, where $p \in \mathbf{P}$ is called an *exception* to generalisation $g \in \mathbf{G}$.

An exception arc directed from p to g indicates that p provides exceptional circumstances under which g may not hold.

Example 10. In the running example, proposition *lie*, which states that Marjan had reason to lie when giving her testimony, provides an exception to evidential generalisation g_6: $tes_3 \rightarrow \neg murder$ in $\mathbf{G}^e$. In Fig. 2a, this is indicated by a curved hyperarc exc: $lie \rightsquigarrow g_6$ in $\mathbf{Exc}$. $\square$

[1] Note that while we only consider ordinary negation in this paper, more general notions of conflicts such as contrariness (see e.g. [30]) are also available.

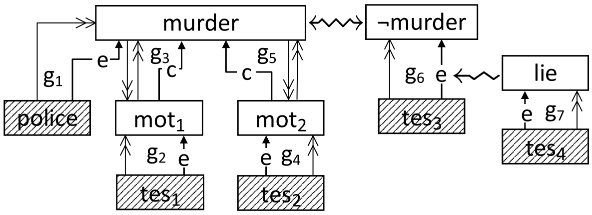


Fig. 4. The IG of Fig. 2a, where evidence $\mathbf{E_p}$ and resulting inference steps ($\twoheadrightarrow$) are also indicated.

4.2. Reading inferences from information graphs

We now define how deductive and abductive inferences can be read from IGs. By itself, a generalisation arc only expresses that the tails together allow us to infer the head in case this generalisation is used in deductive inference, or that the tails together can be inferred from the head in case of abductive inference. Only when considering the available evidence can directionality of inference actually be read from the graph.

Definition 8 *(Evidence set).* Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG. An *evidence set* is a subset $\mathbf{E_p} \subseteq \mathbf{P}$ such that for every $p \in \mathbf{E_p}$ it holds that $\neg p \notin \mathbf{E_p}$.

The restriction that for every $p \in \mathbf{E_p}$ it holds that $\neg p \notin \mathbf{E_p}$ ensures that not both a proposition and its negation are observed. In figures in this paper, nodes in $G_{\mathcal{I}}$ corresponding to elements of $\mathbf{E_p}$ are shaded and all shaded nodes correspond to elements of $\mathbf{E_p}$. We emphasise that various evidence sets $\mathbf{E_p}$ can be used to establish (different) inferences from the same IG.

Example 11. In the running example, the evidence consists of the testimonies and the police report. In Fig. 4, the IG of Fig. 2a is again depicted, with nodes in $\mathbf{E_p} = \{tes_1, tes_2, tes_3, tes_4, police\}$ shaded. □

We now define when we consider configurations of generalisation arcs and evidence to express deductive and abductive inference.

4.2.1. Deductive inference

First, we specify under which conditions we consider a configuration of generalisation arcs and evidence to express deductive inference.

Definition 9 *(Deductive inference).* Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $p_1, \dots, p_n, q \in \mathbf{P}$, with $q \notin \mathbf{E_p}$. Then given $\mathbf{E_p}$, q is *deductively inferred from propositions* $p_1, \dots, p_n$ *using a generalisation* $g\colon \{p_1, \dots, p_n\} \to q$ *in* $\mathbf{G}$, denoted $p_1, \dots, p_n \twoheadrightarrow_g q$, iff $\forall p_i$, $i = 1, \dots, n$:

1. $p_i \in \mathbf{E_p}$, or;
2. p_i is deductively inferred from propositions $r_1, \dots, r_m \in \mathbf{P}$ using a generalisation $g'\colon \{r_1, \dots, r_m\} \to p_i$, where $g' \in \mathbf{G}^e$ if $g \in \mathbf{G}^e$, or;
3. p_i is abductively inferred from a proposition $r \in \mathbf{P}$ using a $g'\colon \{p_i, r_1, \dots, r_m\} \to r$ in $\mathbf{G^c}$, $g \neq g'$, $r_1, \dots, r_m \in \mathbf{P}$ (see Definition 10).

In accordance with our assumptions stated in Sect. 2.1, deductive inference can be performed using both causal and evidential generalisations. The condition $q \notin \mathbf{E_p}$ ensures that deductive inference cannot be performed with a generalisation to infer its consequent in case its consequent is already observed. Deductive inference can only be performed using a generalisation $g \in \mathbf{G}$ to infer its consequent $Head(g)$ from its antecedents $\mathbf{Tails}(g)$ in case every antecedent $p_i \in \mathbf{Tails}(g)$ has been affirmed in that either p_i is observed (i.e. $p_i \in \mathbf{E_p}$), p_i itself is deductively inferred, or p_i is abductively inferred. In correspondence with Pearl's constraint (see Sect. 2.4), we assume in condition 2 that a proposition $q \in \mathbf{P}$ cannot be deductively inferred from $p_1, \dots, p_n \in \mathbf{P}$ using a generalisation $g \in \mathbf{G^e}$ if at least one of its antecedents $p_i \in \mathbf{Tails}(g)$ is deductively inferred using a generalisation $g' \in \mathbf{G^c}$. Condition 3 of Definition 9 is further explained in Sect. 4.2.3, after abductive inference is defined.

Example 12. In the IG of Fig. 4, given $\mathbf{E_p}$ mot_1 and mot_2 are deductively inferred from tes_1 and tes_2 using generalisations g_2 and g_4, respectively, as $tes_1, tes_2 \in \mathbf{E_p}$ (condition 1 of Definition 9). Similarly, $murder$, $\neg murder$ and lie are deductively inferred from $police$, tes_3 and tes_4 using generalisations g_1, g_6 and g_7, respectively, as $police, tes_3, tes_4 \in \mathbf{E_p}$.

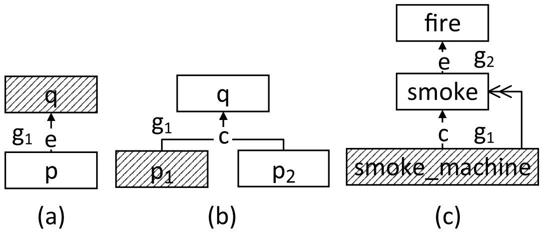


Fig. 5. Examples of IGs illustrating the restrictions put on performing deductive inference within our IG-formalism (a-c).

Proposition *murder* is also deductively inferred from mot_1 and mot_2 using causal generalisations g_3 and g_5, as mot_1 and mot_2 are deductively inferred (condition 2 of Definition 9). This illustrates mixed deductive inference using both evidential and causal generalisations. □

The following example illustrates the restrictions put on performing deductive inference within our IG-formalism.

Example 13. Fig. 5a depicts an example of an IG in which q cannot be deductively inferred from p using g_1, as $Head(g_1) = q \in \mathbf{E_p}$. In Fig. 5b, q cannot be deductively inferred from p_1 and p_2 using g_1, as $p_2 \notin \mathbf{E_p}$ and p_2 is neither deductively nor abductively inferred.

In Fig. 5c, Example 4a from Sect. 2.4 illustrating Pearl's constraint for deductive inference is modelled as an IG. As $smoke\_machine \in \mathbf{E_p}$, *smoke* is deductively inferred from *smoke_machine* using g_1 by condition 1 of Definition 9. *fire* cannot in turn be inferred from *smoke* using g_2, as $g_2 \in \mathbf{G}^e$ and *smoke* is deductively inferred using $g_1 \in \mathbf{G}^c$. □

4.2.2. Abductive inference

Next, we specify under which conditions we consider a configuration of generalisation arcs and evidence to express abductive inference.

Definition 10 *(Abductive inference).* Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $p_1, \ldots, p_n, q \in \mathbf{P}$, with $\{p_1, \ldots, p_n\} \cap \mathbf{E_p} = \emptyset$. Then given $\mathbf{E_p}$, $p_1, \ldots, p_n$ *are abductively inferred from q using a generalisation* $g \colon \{p_1, \ldots, p_n\} \to q$ *in* $\mathbf{G}^c$, denoted $q \twoheadrightarrow_g p_1; \ldots; q \twoheadrightarrow_g p_n$, iff:

1. $q \in \mathbf{E_p}$, or;
2. q is deductively inferred from $r_1, \ldots, r_m \in \mathbf{P}$ using a generalisation $g' \colon \{r_1, \ldots, r_m\} \to q$, $g \neq g'$ (see Definition 9), where $g' \in \mathbf{G} \setminus \mathbf{G}^c$, or;
3. q is abductively inferred from a proposition $r \in \mathbf{P}$ using a generalisation $g' \colon \{q, r_1, \ldots, r_m\} \to r$ in $\mathbf{G}^c$, $r_1, \ldots, r_m \in \mathbf{P}$.

In accordance with our assumptions stated in Sect. 2.2, abduction is modelled using only causal generalisations and not evidential generalisations. The condition $\{p_1, \ldots, p_n\} \cap \mathbf{E_p} = \emptyset$ ensures that abduction cannot be performed with a causal generalisation to infer its antecedents in case at least one of its antecedents is already observed. Furthermore, abduction can only be performed using a generalisation $g \in \mathbf{G}^c$ to infer its antecedents $\mathbf{Tails}(g)$ from its consequent $Head(g)$ in case $Head(g)$ has been affirmed in that either $Head(g)$ is observed (i.e. $Head(g) \in \mathbf{E_p}$), $Head(g)$ is deductively inferred, or $Head(g)$ is itself abductively inferred. In correspondence with Pearl's constraint (see Sect. 2.4), we assume in condition 2 that propositions $p_1, \ldots, p_n \in \mathbf{P}$ cannot be abductively inferred from a proposition $q \in \mathbf{P}$ using a generalisation $g \in \mathbf{G}^c$ if its consequent q is deductively inferred using a generalisation $g' \neq g$, $g' \in \mathbf{G}^c$.

Example 14. In the IG of Fig. 6a, p is abductively inferred from q using generalisation $g_1 \in \mathbf{G}^c$ by condition 2 of Definition 10, as q is deductively inferred from r using generalisation $g_2 \in \mathbf{G}^e$ by condition 1 of Definition 9. In the IG of Fig. 6b, q and r_1 are abductively inferred from r using generalisation $g_3 \colon \{q, r_1\} \to r$ in $\mathbf{G}^c$ by condition 1 of Definition 10, as $r \in \mathbf{E_p}$. Then by condition 3 of Definition 10, p_1 and p_2 are abductively inferred from q using generalisations g_1 and g_2, respectively. □

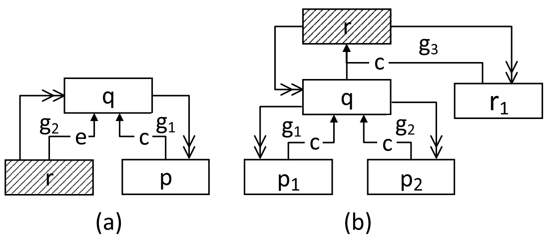


Fig. 6. Examples of IGs illustrating abductive inference (a-b).

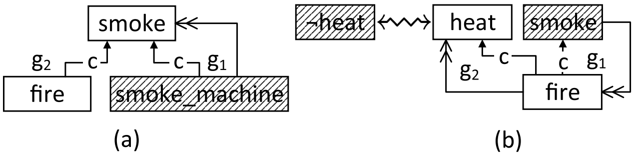


Fig. 7. An IG illustrating Pearl's constraint for mixed deductive-abductive inference (a); an IG illustrating prediction (b).

The following example illustrates that Pearl's constraint for mixed deductive-abductive inference is adhered to (see Sect. 2.4).

Example 15. In Fig. 7a, Example 4b from Sect. 2.4 is modelled as an IG. As *smoke_machine* $\in \mathbf{E_p}$, *smoke* is deductively inferred from *smoke_machine* using g_1. *fire* cannot be inferred from *smoke*, as $g_2 \in \mathbf{G}^C$ and *smoke* is deductively inferred using $g_1 \in \mathbf{G}^C$ (condition 2 of Definition 10). $\square$

4.2.3. Prediction

Our IG-formalism allows for predictive reasoning (deductive inference with causal generalisations, see Sect. 2.1).

Remark 1 *(Prediction).* Let $G_\mathcal{I} = (\mathbf{P}, \mathbf{A}_\mathcal{I})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $g \in \mathbf{G}^C$. Then $Head(g)$ is *predicted* from $\mathbf{Tails}(g)$ iff $Head(g)$ is deductively inferred from $\mathbf{Tails}(g)$.

Example 16. In Fig. 7b, Example 3 from Sect. 2.4 illustrating prediction is modelled as an IG. From *smoke*, *fire* is abductively inferred using g_1, as *smoke* $\in \mathbf{E_p}$. Then *heat* is deductively inferred (or predicted) from *fire* using g_2 (condition 3 of Definition 9). $\square$

In the above example, prediction is performed with g_2 by affirming its antecedent *fire* via abductive inference; besides illustrating prediction, this example thus also illustrates that mixed abductive-deductive inference can be performed within our IG-formalism, as apparent from Definitions 9 and 10.

4.2.4. Ambiguous inference

The conditions under which we consider a configuration of generalisation arcs and evidence to express deductive and abductive inference according to Definitions 9 and 10 are not mutually exclusive. Under specific conditions, both inference types can be established from the same causal generalisation in an IG given the provided evidence; the inference type is, therefore, ambiguous (see Sect. 2.5). The following result follows directly from Definitions 9 and 10.

Remark 2 *(Ambiguous inference).* Let $G_\mathcal{I} = (\mathbf{P}, \mathbf{A}_\mathcal{I})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $g \in \mathbf{G}^C$ with $Head(g) = q$, $\mathbf{Tails}(g) = \{p_1, \ldots, p_n\}$, and $p_1, \ldots, p_n, q \notin \mathbf{E_p}$. Assume that for every $p_1, \ldots, p_n, q$, it holds that it is deductively or abductively inferred. Then q is deductively inferred from $p_1, \ldots, p_n$ and $p_1, \ldots, p_n$ are abductively inferred from q using g.

Example 17. Consider the IG of Fig. 4. Given $\mathbf{E_p}$, *murder* is deductively inferred from *police* using g_1 and mot_1 and mot_2 are deductively inferred from tes_1 and tes_2 using g_2 and g_4, respectively. As *murder*, mot_1, $mot_2 \notin \mathbf{E_p}$, *murder* is deductively inferred from mot_1 and mot_2 and mot_1 and mot_2 are abductively inferred from *murder* using g_3 and g_5, respectively. $\square$

4.2.5. Competing alternative explanations

Finally, we consider how the concept of competing alternative explanations (see Sect. 2.2) is captured within our IG-formalism.

Definition 11 *(Competing alternative explanations).* Let $G_\mathcal{I} = (\mathbf{P}, \mathbf{A}_\mathcal{I})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $g, g' \in \mathbf{G}^C$ with $g \neq g'$, $Head(g) = Head(g') = p$, and possibly $\mathbf{Tails}(g) \cap \mathbf{Tails}(g') \neq \emptyset$. Then given $\mathbf{E_p}$, $\mathbf{Tails}(g)$ is considered to be *in competition* with $\mathbf{Tails}(g')$ for the same effect expressed by p in case $\mathbf{Tails}(g)$ and $\mathbf{Tails}(g')$ are abductively inferred from p given $\mathbf{E_p}$ using g and g', respectively, and p can neither be deductively inferred from $\mathbf{Tails}(g)$ nor from $\mathbf{Tails}(g')$ given $\mathbf{E_p}$ using g or g', respectively.

The condition that $g, g' \in \mathbf{G}^C$ with $g \neq g'$, $Head(g) = Head(g') = p$ implies that every pair of propositions p_i, q_j for $p_i \in \mathbf{Tails}(g)$, $q_j \in \mathbf{Tails}(g')$ are considered alternative causes of p by condition 2a of Definition 3. The condition that only abduction and not deduction is performed with g and g' implies that the inference type for neither g nor g' is ambiguous (see Remark 2). The above definition captures competition between *sets* of propositions $\mathbf{Tails}(g)$ and $\mathbf{Tails}(g')$, as these sets are abductively inferred from p using g and g', respectively. More specifically, individual propositions in $\mathbf{Tails}(g)$ are not in competition with individual propositions in $\mathbf{Tails}(g')$ in case separate causal generalisations $g_i: p_i \rightarrow p$ and $g'_j: q_j \rightarrow p$

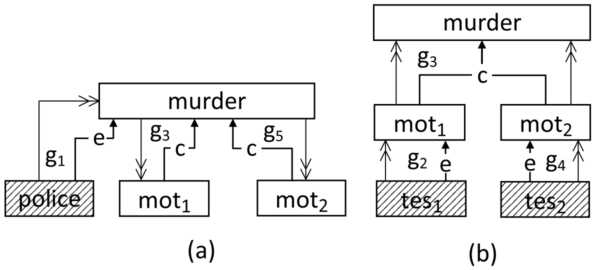


Fig. 8. Adjustment to the IG of Fig. 2a involving two competing alternative explanations mot_1 and mot_2 for *murder* (a); the IG of Fig. 2b with evidence $\mathbf{E_p}$ and resulting inference steps now indicated, involving two non-competing alternative explanations mot_1 and mot_2 for *murder* (b).

for $p_i \in \mathbf{Tails}(g)$, $q_j \in \mathbf{Tails}(g')$ are not provided. In case a causal generalisation arc has multiple tails, we assume that these tails are not in competition *among themselves*, as the generalisation expresses that only the tails together allow us to infer the head.

Example 18. Consider Fig. 8a, which depicts an adjustment to the IG of Fig. 2a. Given $\mathbf{E_p} = \{police\}$, propositions mot_1 and mot_2 are abductively inferred from *murder* using g_3 and g_5, respectively, as *murder* is deductively inferred from *police* using g_1. Furthermore, *murder* can neither be deductively inferred from mot_1 nor from mot_2 using g_3 or g_5, respectively. Therefore, mot_1 and mot_2 are in competition for common effect *murder*.

In Fig. 8b, the IG of Fig. 2b is again depicted, where evidence $\mathbf{E_p} = \{tes_1, tes_2\}$ and resulting inferences are also indicated. In this IG, *murder* is deductively inferred from $\{mot_1, mot_2\}$ given $\mathbf{E_p}$ using $g_3 \colon \{mot_1, mot_2\} \rightarrow murder$ in $\mathbf{G}^c$; therefore, mot_1 and mot_2 are *not* in competition for *murder*. □

5. Bayesian networks

In this section, Bayesian networks (BNs) [20] are reviewed. A BN compactly represents a joint probability distribution $\Pr(\mathbf{V})$ over a finite set of discrete random variables $\mathbf{V}$; in this paper we assume all variables to be Boolean, where we write v to denote $\mathrm{V} = true$ and $\neg v$ to denote $\mathrm{V} = false$. Formally, a BN is defined as follows:

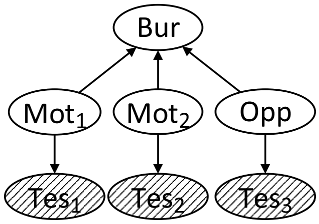
Definition 12 *(Bayesian network).* A *Bayesian network* (BN) is a pair $(G_{\mathcal{B}}, \Pr)$, where $G_{\mathcal{B}}$ is a directed acyclic graph (DAG) $(\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ over nodes $\mathbf{V}$ representing random variables.[2] $\mathbf{A}_{\mathcal{B}} \subseteq \mathbf{V} \times \mathbf{V}$ is a set of directed arcs $\mathrm{V}_i \rightarrow \mathrm{V}_j$ from parent $\mathrm{V}_i \in \mathbf{V}$ to child $\mathrm{V}_j \in \mathbf{V}$, where $\mathbf{Par}(\mathrm{V})$ denotes the set of parents of V and $\mathbf{Ch}(\mathrm{V})$ denotes the set of children of V. Pr is a probability function which specifies for each variable $\mathrm{V} \in \mathbf{V}$ a conditional probability table (CPT). This CPT describes the conditional probability distributions $\Pr(\mathrm{V} \mid x)$ for each possible joint value combination x for $\mathbf{Par}(\mathrm{V})$.

The reflexive, transitive closures of V under the parent and child relations are denoted by $\mathbf{Par}^*(\mathrm{V})$ and $\mathbf{Ch}^*(\mathrm{V})$, respectively, where nodes in $\mathbf{Par}^*(\mathrm{V})$ are called *ancestors* of V and nodes in $\mathbf{Ch}^*(\mathrm{V})$ are called *descendants* of V.

A BN is generally used for *probabilistic inference* [20], that is, calculating any prior or posterior distribution over the variables represented in the network. Posterior distributions are obtained by *instantiating* one or more variables $\mathbf{E_V} \subseteq \mathbf{V}$ in that they are set to a specific value. Instantiations are also called evidence. The inference algorithms associated with the BN-formalism provide for computing probabilities of interest and for processing evidence; these algorithms constitute the basic building blocks for reasoning with knowledge represented in the formalism. As in the current paper the focus lies on the knowledge that is represented by a BN by means of its graphical structure $G_{\mathcal{B}}$ and its probability function Pr, algorithms for probabilistic inference are not further discussed.

Example 19. An example of a BN graph and one of its CPTs is depicted in Fig. 9, where ovals represent nodes and instantiated nodes are shaded. In this BN graph, we are interested in whether a given suspect committed a burglary (Bur). This node is connected by arcs to nodes Mot_1, Mot_2 and Opp, which describe whether the suspect had motive(s) and opportunity to commit the burglary. In turn, nodes Mot_1, Mot_2 and Opp are connected to instantiated nodes Tes_1, Tes_2 and Tes_3, which capture the testimonies provided to these claims. □

[2] There is a one-to-one correspondence between nodes and variables in BNs. Throughout this paper, the terms ‘node’ and ‘variable’ are, therefore, used interchangeably.



(a)

			Bur	
Mot_1	Mot_2	Opp	t	f
tt	t		0.90	0.10
ttf			0.20	0.80
tf	t		0.70	0.30
tf	f		0.10	0.90
ft	t		0.80	0.20
ft	f		0.10	0.90
fft			0.05	0.95
ff	f		0.01	0.99

(b)

Fig. 9. An example of a BN (a); CPT for node Bur (b), where Mot_1 and Mot_2 exhibit a negative product synergy wrt value Bur = *true* in presence of uninstantiated parent Opp.

5.1. Bayesian network graphs

The BN graph $G_{\mathcal{B}}$ encodes the probabilistic independence relation among its variables by means of the notion of d-separation, which is defined by the notions of blocked and active chains. In the following, let $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ be a BN graph.

Definition 13 *(Chain).* A *chain* $c = (V_1, A_1, V_2, \ldots, A_{n-1}, V_n)$ is a sequence of distinct nodes $V_1, \ldots, V_n \in \mathbf{V}$ and arcs $A_1, \ldots, A_{n-1} \in \mathbf{A}_{\mathcal{B}}$ such that for every $A_i, 1 \le i < n$, it holds that either $A_i \equiv V_i \rightarrow V_{i+1}$ or $A_i \equiv V_{i+1} \rightarrow V_i$.

Definition 14 *(Head-to-head node).* A node $V \in \mathbf{V}$ is called a *head-to-head node* on a chain c in $G_{\mathcal{B}}$ if it has two incoming arcs on c.

Definition 15 *(Blocked chain).* A chain c between nodes $V_1 \in \mathbf{V}$ and $V_2 \in \mathbf{V}$ in $G_{\mathcal{B}}$ is *blocked* by a (possibly empty) set of instantiated nodes iff it includes a node $V \notin \{V_1, V_2\}$ such that either:

- V is an uninstantiated head-to-head node on c without instantiated descendants, or;
- V is instantiated and has at most one incoming arc on c.

A chain that is not blocked by the evidence is called *active*.

Definition 16 *(d-separation).* Two sets of nodes $\mathbf{V_1} \subseteq \mathbf{V}$ and $\mathbf{V_2} \subseteq \mathbf{V}$ are *d-separated* by a set of nodes $\mathbf{Z} \subseteq \mathbf{V}$ iff there exist no active chains between any node in $\mathbf{V_1}$ and any node in $\mathbf{V_2}$ given instantiations for nodes $\mathbf{Z}$.

If $\mathbf{V_1}$ and $\mathbf{V_2}$ are d-separated given instantiations for $\mathbf{Z} \subseteq \mathbf{V}$, then their corresponding variables are considered conditionally independent given $\mathbf{Z}$.

Example 20. In Fig. 9, given the evidence for $\mathbf{Z} = \{Tes_1, Tes_2, Tes_3\}$ all chains between Mot_1 and Mot_2 are blocked, as Bur is an uninstantiated head-to-head node without instantiated descendants on chain (Mot_1, $Mot_1 \rightarrow$ Bur, Bur, $Mot_2 \rightarrow$ Bur, Mot_2); hence, Mot_1 and Mot_2 are considered conditionally independent given the evidence for $\mathbf{Z}$. $\square$

Finally, we review the following concept from graph theory.

Definition 17 *(Weakly connected component).* Let $G = (\mathbf{V}, \mathbf{A})$ be a directed graph and let $C = (\mathbf{V}^C, \mathbf{A}^C)$ with $\mathbf{V}^C \subseteq \mathbf{V}$ and $\mathbf{A}^C \subseteq (\mathbf{V}^C \times \mathbf{V}^C) \cap \mathbf{A}$ be a sub-graph of G. Then C is a *weakly connected component* of G iff:

1. For every pair of nodes $V_1, V_2 \in \mathbf{V}^C$, there exists a chain between V_1 and V_2 in C;
2. C is a maximal sub-graph of G for which property 1 holds.

5.2. Intercausal interactions and qualitative probabilistic constraints

Next, we review the concepts of intercausal interactions and qualitative probabilistic constraints. In case a head-to-head node or one of its descendants in a BN graph is instantiated, an active chain is *induced* between the parents of the head-to-head node, allowing for *intercausal interactions*.[3] If one of the parents is *true*, then the probability of another parent being *true* as well may change, depending on the synergistic effect modelled in the CPT for the head-to-head node. In case the probability that one of the other parents is *true* decreases, this is called the *'explaining away'* effect [13]. For Boolean nodes, we will generally assume an ordering *true* > *false* on its values unless specified otherwise. In case this ordering is reversed, then the occurrences of these two values need to be interchanged in the equations appearing in Definitions 18 and 20. To achieve the explaining away effect between two parents V_1 and V_3 of V_2 for value v_2, the CPT for V_2 needs to be constrained such that V_1 and V_3 exhibit a *negative product synergy wrt* v_2. First, we review the concept of product synergy I [13], which captures the special case in which all other parents of V_2 are instantiated.

Definition 18 *(Product synergy I).* Let $\mathbf{B} = (G_{\mathcal{B}}, \Pr)$ be a BN and let $\mathrm{V}_1, \mathrm{V}_3 \in \mathbf{V}$ be parents of $\mathrm{V}_2 \in \mathbf{V}$ in $G_{\mathcal{B}}$. Let $\mathbf{X} = \mathbf{Par}(\mathrm{V}_2) \setminus \{\mathrm{V}_1, \mathrm{V}_3\}$ and let x be the combination of observed values for $\mathbf{X}$. Then V_1 and V_3 exhibit a *negative product synergy* wrt v_2, written $\mathbf{X}^-(\{\mathrm{V}_1, \mathrm{V}_3\}, v_2)$, iff

$$\Pr(v_2 \mid v_1, v_3, x) \cdot \Pr(v_2 \mid \neg v_1, \neg v_3, x) \leq \Pr(v_2 \mid v_1, \neg v_3, x) \cdot \Pr(v_2 \mid \neg v_1, v_3, x)$$

In case $\mathbf{X} = \emptyset$, then this equation simplifies by leaving out every occurrence of x. V_1 and V_3 exhibit a *zero product synergy* wrt v_2, written $\mathbf{X}^0(\{\mathrm{V}_1, \mathrm{V}_3\}, v_2)$, in case $\leq$ in the above equation is replaced by $=$. In this case, no direct intercausal interaction effect exists between parents V_1 and V_3 for value v_2 of V_2. V_1 and V_3 exhibit a *positive product synergy* wrt v_2, written $\mathbf{X}^+(\{\mathrm{V}_1, \mathrm{V}_3\}, v_2)$, in case $\leq$ is replaced by $\geq$ in the above equation. In this case, the joint occurrence of the causes may be a more likely explanation of the common effect than would either of them considered individually.

Next, the case is considered in which $\mathbf{X} \neq \emptyset$ is not instantiated to a combination of values. First, we review the concept of *matrix half negative semi-definiteness*.

Definition 19 *(Half negative semi-definite matrix).* Let M be a square $n \times n$ matrix, $n \geq 1$, and let x be any non-negative vector x of n elements. Then M is called *half negative semi-definite* iff $x^T M x \leq 0$.

Similarly, a square matrix M is called *half positive semi-definite* iff $x^T M x \geq 0$ for any non-negative vector x of n elements. We now provide the definition of extended product synergy, termed *product synergy II* [13].

Definition 20 *(Product synergy II).* Let $\mathbf{B} = (G_{\mathcal{B}}, \Pr)$ be a BN and let V_1, $\mathrm{V}_3 \in \mathbf{V}$ be parents of $\mathrm{V}_2 \in \mathbf{V}$ in $G_{\mathcal{B}}$. Let $\mathbf{X} = \mathbf{Par}(\mathrm{V}_2) \setminus \{\mathrm{V}_1, \mathrm{V}_3\}$. Let n denote the number of possible combinations of values for $\mathbf{X}$. Then V_1 and V_3 exhibit a *negative product synergy* wrt v_2 iff the $n \times n$ matrix M with elements $M_{ij} = \Pr(v_2 \mid v_1, v_3, x_i) \cdot \Pr(v_2 \mid \neg v_1, \neg v_3, x_j) - \Pr(v_2 \mid v_1, \neg v_3, x_i) \cdot \Pr(v_2 \mid \neg v_1, v_3, x_j)$ is half negative semi-definite for all combinations of values x_i and x_j for $\mathbf{X}$.

For a positive or zero product synergy, the matrix M has to be half positive semi-definite or zero, respectively. Note that product synergy I is a special case of product synergy II; hence, in referring to the general concept of product synergy throughout this article, we are referring to product synergy II.

Example 21. Consider the BN of Fig. 9. The entries of the CPT of Fig. 9b are chosen such that Mot_1 and Mot_2 exhibit a negative product synergy wrt value Bur = *true* in presence of uninstantiated parent Opp. Specifically, the 2×2 matrix M consisting of the following four elements is half negative semi-definite:

$$M_{11} = 0.9 \cdot 0.05 - 0.7 \cdot 0.8 = -0.515; \; M_{12} = 0.9 \cdot 0.01 - 0.7 \cdot 0.1 = -0.061$$
$$M_{21} = 0.2 \cdot 0.05 - 0.1 \cdot 0.8 = -0.070; \; M_{22} = 0.2 \cdot 0.01 - 0.1 \cdot 0.1 = -0.008 \quad \square$$

5.3. BN construction

BN construction is typically an iterative process. After constructing an initial BN graph, it should be verified that it is acyclic and that it correctly captures the (conditional) independencies. If the graph does not yet exhibit these properties, arcs should be reversed, added or removed by the BN modeller in consultation with the domain expert. We call this the *'graph validation step'*. Related research on BN graph construction is reviewed in Sect. 9.2.

The (conditional) probabilities of the BN are elicited in a separate quantification step. In the current paper, the focus lies on deriving the graphical structure of BNs and not on deriving the modelled probability distribution, although in some cases qualitative constraints on the (conditional) probabilities of the BN under construction in the form of product synergies are derived that can subsequently be used in the quantification step.

[3] We note that, while the term 'intercausal interactions' is used, these interactions can also occur regardless of the type of relation between parents and child.

6. Constructing Bayesian network graphs from information graphs

Based on our IG-formalism, we now propose a structured approach for automatically constructing a directed BN graph from an IG. In our approach, we focus on exploiting the knowledge captured in an IG to constrain the graphical structure of the BN and the conditional independence relation it encodes by means of the d-separation criterion, as well as constraining its probability function by means of product synergies.

Our IG-formalism serves as an intermediary formalism between analyses performed using informal reasoning tools and BNs. We expect direct IG construction to be more straightforward than direct BN construction for domain experts unfamiliar with the BN-formalism, a claim we intend to empirically evaluate in our future work. We believe this to be a plausible assumption, however, among other things due to the fact that the arcs of a BN are easily misinterpreted by domain experts unfamiliar with BNs as non-symmetric relations of cause and effect instead of collectively encoding an independence relation [12], making manual BN construction a difficult and error-prone process (see also [18]). Moreover, it is justified to assume that information regarding causality is present in the domain expert's original analysis (see [2,6]), and in manual BN graph construction, conditional independencies are typically not directly elicited, but instead the *notion of causality* is commonly used as a guiding principle [16,20].

In IGs, causality information is made explicit by means of causal and evidential generalisations and can thus be directly used in BN graph construction. Whereas the ultimate goal of our approach is to facilitate domain experts in constructing BNs that can be used to evaluate their problems in a probabilistic manner, our proposed approach only serves for constructing an initial BN graph and for deriving qualitative constraints on the probabilities of the BN under construction. More specifically, as IGs only express qualitative and not quantitative (probabilistic) information, our BN construction approach can only serve for constructing a partially specified initial BN. Moreover, the qualitative probabilistic constraints that are derived from an IG given the evidence are generally only a subset of those required for the specification of a QPN [33] (see also Sect. 9.3). Hence, initial BNs constructed by our approach are only partially specified and cannot be directly used for probabilistic inference. The derived constraints may serve as input for a subsequent elicitation procedure for obtaining a fully specified QPN or BN for (qualitative) probabilistic inference.

In Sects. 6.1 and 6.2 we motivate the steps of our approach for automatically constructing an initial BN graph from an IG; the approach itself is presented in Sect. 6.3. In Sect. 6.4 we then explain and illustrate the steps of our approach with several examples.

6.1. Extracting information from an IG

First, we consider the graphical structure of the BN. For constructing a BN graph from an IG, the IG's structure is used, specifically the generalisations, exceptions and negations expressed in the graph.

Information in proposition nodes For every proposition $p \in \mathbf{P}$ in an IG, we propose to form a single BN node in $\mathbf{V}$ describing both values p and $\neg p$, as captured by step 1 of our approach. By this step, two propositions $p, -p \in \mathbf{P}$ involved in negation are captured as two mutually exclusive values of the same node. Negation arcs present in an IG can thus be disregarded in BN construction, as such arcs are drawn between a pair $p, q \in \mathbf{P}$ iff $q = -p$.

Information in causal and evidential generalisations In the manual construction of BN graphs, arcs are typically directed using the notion of causality as a guiding principle [16,20]. Specifically, if the domain expert indicates that p or $\neg p$ typically causes q or $\neg q$, then the arc is set from node P to node Q. By following this heuristic, causes form a head-to-head connection in the node corresponding to their common effect. As such, possible interactions between causes, for example due to the fact that they could be in competition, can be directly captured in the CPT for this node. Hence, we propose to use the same heuristic in automatically directing arcs, where we exploit causality information explicitly expressed in an IG by means of causal and evidential generalisations. Specifically, arcs in the BN graph are set in the same direction as generalisation arcs in $\mathbf{G}^{\mathrm{c}}$ and in the opposite direction for generalisation arcs in $\mathbf{G}^{\mathrm{e}}$. This is captured by step 2 of our approach.

Information in exceptions Arcs in **Exc** denote exceptions to generalisations. For instance, if a generalisation is in the evidential direction, then an exception suggests an alternative cause for the same effect. Exceptions to causal generalisations do not suggest alternative causes for the same effect, but do possibly interact with them (examples are provided in Sect. 6.4.2). Accordingly, we propose to enable capturing possible interactions between an exception and a generalisation arc, if any, in the CPTs for head-to-head nodes formed in the BN graph. This is captured by step 3 of our approach.

6.2. Exploiting induced inferences expressed by IGs

By itself, a generalisation arc only captures knowledge about the world in conditional form; only when considering the available evidence $\mathbf{E_p}$ in the IG can directionality of inference be read from the graph. In comparison, from a BN graph we can read the chains between nodes that are active given the evidence and will be exploited to propagate the evidence upon probabilistic inference. In our approach, we want to ensure that the sequences of propositions that can be iteratively inferred from each other given $\mathbf{E_p}$ in an IG are captured in the BN graph by means of active chains given the available evidence for $\mathbf{E_V} \subseteq \mathbf{V}$ corresponding to $\mathbf{E_p} \subseteq \mathbf{P}$. In Sect. 7.2, we formally prove that BN graphs constructed by our approach indeed allow reasoning patterns similar to the sequences of propositions that can be iteratively inferred from each other given $\mathbf{E_p}$ in the original IG.

Exploiting competing alternative explanations Probabilistic constraints on the BN under construction are derived by considering the inferences that can be read from an IG given $\mathbf{E_p}$. In case the tails of two causal generalisations are competing alternative explanations for the common effect expressed by the head given $\mathbf{E_p}$ (see Definition 11), we propose to constrain the CPT for the variable corresponding to the head such that the explaining away effect can occur between the variables corresponding to the tails of the generalisations, as captured by step 5a. In case abductive inference is performed with a generalisation given $\mathbf{E_p}$, then the tails are not in competition *among themselves* and the explaining away effect should not occur, as captured by step 5b. Similarly, the tails of a generalisation are not in competition among themselves if deductive inference is performed, which is captured by the same step.

We note that various evidence sets $\mathbf{E_p}$ can be used to establish inferences from the same IG, and thus that, depending on $\mathbf{E_p}$, different probabilistic constraints may be derived on the BN under construction. The structure of the BN does not depend on $\mathbf{E_p}$, as the IG's structure is used in BN graph construction and not the IG's inferences.

Exploiting interactions between exceptions and generalisations The presence of an exception to a generalisation g weakens an inference step performed with g. Depending on whether deductive or abductive inference is performed with g given $\mathbf{E_p}$, different probabilistic constraints are derived, as captured by step 6 of our approach.

6.3. The approach

In this subsection, we present the steps of our approach. Let $\mathrm{Var}\colon \mathbf{P} \to \mathbf{V}$ be an operator mapping every proposition p or $\neg p \in \mathbf{P}$ in an IG to a BN node $\mathrm{Var}(p) = \mathrm{Var}(\neg p) \in \mathbf{V}$ describing values p and $\neg p$. For an IG $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$, a BN graph $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ is constructed as follows:

1) $\forall p, \neg p \in \mathbf{P}$, include $\mathrm{Var}(p)$ in $\mathbf{V}$; if p or $\neg p \in \mathbf{E_p}$, also include $\mathrm{Var}(p)$ in $\mathbf{E_V}$.
2) For every generalisation arc $g\colon \{p_1, \ldots, p_n\} \to p$:
   2a) If $g \in \mathbf{G}^{\mathrm{e}}$, include $\mathrm{Var}(p) \to \mathrm{Var}(p_i)$, $i = 1, \ldots, n$ in $\mathbf{A}_{\mathcal{B}}$.
   2b) If $g \in \mathbf{G}^{\mathrm{c}}$, include $\mathrm{Var}(p_i) \to \mathrm{Var}(p)$, $i = 1, \ldots, n$ in $\mathbf{A}_{\mathcal{B}}$.

3) For every exception arc $exc\colon p \rightsquigarrow g$ in **Exc** with $g\colon \{q_1, \ldots, q_n\} \to q$:
   3a) If $g \in \mathbf{G}^{\mathrm{e}}$, include $\mathrm{Var}(p) \to \mathrm{Var}(q_i)$, $i = 1, \ldots, n$ in $\mathbf{A}_{\mathcal{B}}$.
   3b) If $g \in \mathbf{G}^{\mathrm{c}}$, include $\mathrm{Var}(p) \to \mathrm{Var}(q)$ in $\mathbf{A}_{\mathcal{B}}$.

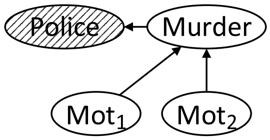
While our approach exploits the domain knowledge captured in the IG in constructing an initial BN graph, the IG may lack information needed to prevent cycles and unwarranted (in)dependencies in the obtained BN graph; hence, the following manual validation step should be performed by the BN modeller in consultation with the domain expert. We note that this type of validation is standard in BN construction, especially in data-poor domains (see Sect. 5.3):

4) Manually verify the properties of the constructed graph $G_{\mathcal{B}}$ by applying the standard graph validation step (see Sect. 5.3).

We define the following probabilistic constraints on the BN under construction:

5) For every generalisation arc $g\colon \mathbf{P_1} \to q$ in $\mathbf{G}$, $\mathbf{P_1} = \{p_1, \ldots, p_n\} \subseteq \mathbf{P}$:
   5a) $\forall g'\colon \mathbf{Q} \to q$ in $\mathbf{G}$, $\mathbf{Q} = \{q_1, \ldots, q_m\} \subseteq \mathbf{P}$, $g \neq g'$ such that both $g, g' \in \mathbf{G}^{\mathrm{c}}$ and for which, given $\mathbf{E_p}$, $\mathbf{P_1}$ and $\mathbf{Q}$ are competing alternative explanations for the common effect expressed by q (see Definition 11), constrain the CPT for $\mathrm{Var}(q)$ such that $\mathbf{X}^{-}(\{\mathrm{Var}(p_i), \mathrm{Var}(q_j)\}, q)$ for $p_i \in \mathbf{P_1} \setminus \mathbf{Q}$, $q_j \in \mathbf{Q} \setminus \mathbf{P_1}$.
   5b) If g is used to perform inference given $\mathbf{E_p}$, constrain the CPT for $\mathrm{Var}(q)$ such that $\mathbf{X}^{\delta}(\{\mathrm{Var}(p_i), \mathrm{Var}(p_j)\}, q)$ with $\delta \neq -$, $p_i, p_j \in \mathbf{P_1}$, $p_i \neq p_j$.

6) For every $exc\colon p \rightsquigarrow g$ in **Exc** with $p \in \mathbf{P}$ and $g\colon \{q_1, \ldots, q_n\} \to q$ in $\mathbf{G}$:
   6a) If $g \in \mathbf{G}^{\mathrm{e}}$ and q is deductively inferred from $q_1, \ldots, q_n$ given $\mathbf{E_p}$ using g, constrain the CPT for $\mathrm{Var}(q_i)$ such that $\mathbf{X}^{-}(\{\mathrm{Var}(p), \mathrm{Var}(q)\}, q_i)$, $i = 1, \ldots, n$. If in addition $\exists exc'\colon p' \rightsquigarrow g$ in **Exc**, further constrain the CPT for $\mathrm{Var}(q_i)$ such that $\mathbf{X}^{-}(\{\mathrm{Var}(p), \mathrm{Var}(p')\}, q_i)$, $i = 1, \ldots, n$.
   6b) If $g \in \mathbf{G}^{\mathrm{c}}$ and q is deductively inferred from $q_1, \ldots, q_n$ given $\mathbf{E_p}$ using g, constrain the CPT for $\mathrm{Var}(q)$ such that $\Pr(q \mid p, q_1, \ldots, q_n) < \Pr(q \mid \neg p, q_1, \ldots, q_n)$.
   6c) If $g \in \mathbf{G}^{\mathrm{c}}$ and $q_1, \ldots, q_n$ are abductively inferred from q given $\mathbf{E_p}$ using g, constrain the probabilities of the BN such that $\Pr(q_i \mid p, q) < \Pr(q_i \mid \neg p, q)$, $i = 1, \ldots, n$.



(a)

		Murder	
Mot_1	Mot_2	t	f
tt		0.4	0.6
tf		0.6	0.4
ft		0.5	0.5
ff		0.1	0.9

(b)

Fig. 10. BN graph constructed from the IG of Fig. 8a by our approach (a); a possible CPT for node Murder (b).

We reiterate that the initially constructed BN by our approach should always be verified by the BN modeller in consultation with the domain expert, which includes verifying the derived probabilistic constraints. After this verification step, the derived constraints can be used in subsequent probability assessment, thereby partially simplifying it. In particular, since we are considering BN construction in data-poor domains the required conditional probabilities will often need to be elicited from domain experts, where it can be monitored whether the assessed conditional probabilities satisfy the derived probabilistic constraints.

We note that the above probabilistic constraints concern intercausal interactions between individual nodes and not sets, as to the best of our knowledge no approaches have been proposed in the literature that allow for capturing interactions between sets of parents of a node. The type of competition between sets of nodes in an IG as captured by Definition 11 can, therefore, not be straightforwardly captured between variables in a corresponding BN; instead, in step 5a we propose to constrain the CPT for $\text{Var}(q)$ such that $\mathbf{X}^-(\{\text{Var}(p_i), \text{Var}(q_j)\}, q)$ for *pairs* of propositions $p_i \in \mathbf{P_1} \setminus \mathbf{Q}$, $q_j \in \mathbf{Q} \setminus \mathbf{P_1}$, where the intersection of $\mathbf{P_1}$ and $\mathbf{Q}$ is not considered. Similarly, in step 6a interactions between pairs of nodes and not sets are considered. In our future work, we intend to investigate whether the concept of product synergy can be extended to sets of nodes.

6.4. *Explanation and illustration of the steps of the approach*

In this section, we explain and illustrate the steps of our approach through our running example, introduced in Sect. 3. In Sect. 6.4.1 we illustrate that steps $1-2$ of our approach suffice for constructing BN graphs from restricted IGs not including exception arcs, where the CPTs of the BN under construction should be constrained according to step 5. In Sect. 6.4.2 we then illustrate that the BN under construction needs to be further constrained in case exception arcs are present in the IG; this is accounted for in steps 3 and 6 of our approach.

6.4.1. *Explanation and illustration of steps* $1-2$ *and* 5

First, we explain and illustrate the main idea behind our approach by applying it to the IG depicted in Fig. 8a.

***Steps* 1 − 2** The first step is to capture every proposition in $G_{\mathcal{I}}$ and its negation as two mutually exclusive values of the same BN node in $G_{\mathcal{B}}$. In steps 2a and 2b, arcs in the BN graph are directed using the notion of causality in that for every $g \in \mathbf{G}^{\text{c}}$, arcs in the BN graph are directed from nodes corresponding to $\mathbf{Tails}(g)$ to $\text{Var}(Head(g))$, and vice versa for $g \in \mathbf{G}^{\text{e}}$. This formalises the approach typically taken in the manual construction of BN graphs, namely that of setting arcs in the causal direction as a guiding principle [16,20]. The resulting BN graph is depicted in Fig. 10a.

Step 5a The inferences that can be read from an IG given the evidence allow us to derive constraints on the CPTs of the BN. In the IG of Fig. 8a, given $\mathbf{E_p} = \{police\}$ propositions mot_1 and mot_2 are abductively inferred from *murder* using g_3 and g_5, respectively, as given $\mathbf{E_p}$ *murder* is deductively inferred from *police* using g_1. Therefore, mot_1 and mot_2 are competing alternative explanations for common effect *murder* in that accepting one explanation will diminish our belief in the other (see Definition 11). We propose to link this type of intercausal interaction in IGs to the explaining away effect in BNs. Specifically, as proposed in step 5a of our approach, the CPT for Murder should be constrained such that $\mathbf{X}^-(\{\text{Mot}_1, \text{Mot}_2\}, murder)$. Note that the IG only informs us that there should be a negative product synergy between Mot_1 and Mot_2 wrt value Murder $= true$; it does not inform us whether a product synergy should also be exhibited between these variables wrt value Murder $= false$, as proposition $\neg murder$ does not appear in the IG. Fig. 10b depicts a possible CPT for Murder, where $\mathbf{X}^-(\{\text{Mot}_1, \text{Mot}_2\}, murder)$ as $0.4 \cdot 0.1 \leq 0.6 \cdot 0.5$. However, as $0.6 \cdot 0.9 \geq 0.4 \cdot 0.5$, it also holds that $\mathbf{X}^+(\{\text{Mot}_1, \text{Mot}_2\}, \neg murder)$. Care should be taken, therefore, in eliciting the involved probabilities, as it may be undesirable that a positive product synergy for value $\neg murder$ is exhibited.

By following steps 2a and 2b of our approach, causes automatically form a head-to-head connection in the node corresponding to their common effect for any given IG; interactions between causes in an IG, for instance because they are competing alternative explanations for the common effect, can, therefore, always be directly captured in the CPT for the

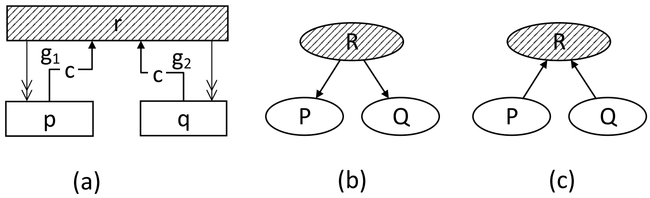


Fig. 11. Example of an IG (a); the BN graph constructed by directing arcs according to the *inferences* that can be read from this IG given $\mathbf{E_p}$ (b); the BN graph constructed by directing arcs according to the *generalisations* in the IG (c).

node corresponding to the common effect. We note that directing arcs in the BN graph in the same direction as the *inferences* that can be read from an IG given the evidence would lead to undesirable results. Consider the IG depicted in Fig. 11a. By directing arcs according to the inferences that can be read from this IG given $\mathbf{E_p}$, the BN graph of Fig. 11b is constructed. In the IG of Fig. 11a, p and q are competing alternative explanations for common effect r given $\mathbf{E_p}$; however, this competition cannot be directly captured in the CPT for node R in the BN graph of Fig. 11b as a divergent connection is formed. Moreover, all chains between P and Q are blocked given $\mathbf{E_V} = \{\mathrm{R}\}$; hence, interactions between causes expressed in an IG cannot always be captured by directing arcs in a corresponding BN graph according to the induced inferences in an IG.

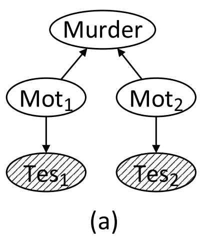
Step 5b Next, consider the IG of Fig. 8b. Given $\mathbf{E_p}$, *murder* is deductively inferred from mot_1 and mot_2 using g_3; therefore, mot_1 and mot_2 are not competing alternative explanations for *murder* in this IG. By following steps $1-2$ of our approach, the BN graph of Fig. 12a is constructed. As mot_1 and mot_2 are not competing alternative explanations for *murder* in this example, we need to assure that the explaining away effect cannot occur between Mot_1 and Mot_2 for value Murder $= true$. This can be achieved by constraining the CPT for Murder such that $\mathbf{X}^{\delta}(\{\mathrm{Mot}_1, \mathrm{Mot}_2\}, murder)$ for $\delta \neq -$, as captured by step 5b of our approach. This is a relaxation of our previously proposed solution [39], in which we proposed to constrain the CPT for the node corresponding to the common effect such that a zero product synergy is exhibited wrt the indicated value in the IG. Specifically, we now also allow that a positive product synergy is exhibited; what counts is that no negative product synergy is exhibited between Mot_1 and Mot_2 for value Murder $= true$, as mot_1 and mot_2 are not competing alternative explanations for the common effect. Fig. 12b depicts a possible CPT for Murder, where $\mathbf{X}^{+}(\{\mathrm{Mot}_1, \mathrm{Mot}_2\}, murder)$ as $0.8 \cdot 0.1 \geq 0.2 \cdot 0.2$.

6.4.2. *Explanation and illustration of steps* 3 *and* 6

Next, IGs including exception arcs are considered.

Step 3a In Fig. 13a, an example of an IG is depicted in which exceptions to both an evidential and a causal generalisation are provided. Proposition *lie*, which states that Marjan had reason to lie when giving her testimony, provides an exception to the evidential generalisation $tes_3 \rightarrow \neg murder$. Since tes_3 is either the result of Marjan truly not committing the murder or due to a lie, $\neg murder$ and *lie* can be seen as competing alternative explanations for Marjan's testimony. Generally, exceptions to an evidential generalisation can be considered competing alternative explanations for the common effects expressed by the antecedents of the generalisation. We therefore propose to enable capturing such interactions between an exception and an evidential generalisation by forming head-to-head nodes in the nodes corresponding to the tails of the generalisation arc. By step 2a of our approach, the BN graph under construction includes arc Murder $\rightarrow$ Tes_3. A head-to-head node can, therefore, be formed in node Tes_3 by adding additional arc Lie $\rightarrow$ Tes_3 to the BN graph; this is captured by step 3a of our approach.

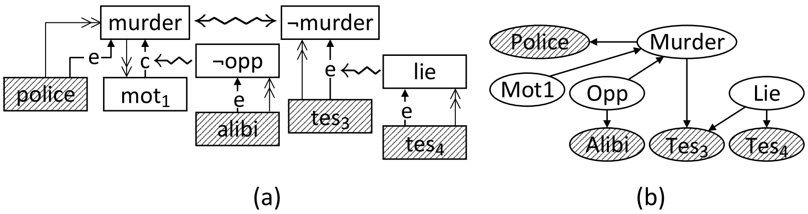
Step 6a Given $\mathbf{E_p} = \{police, alibi, tes_3, tes_4\}$, $\neg murder$ is deductively inferred from tes_3. As proposition *lie* provides an exception to the generalisation used in performing this inference step and thereby weakens the inference, we propose to constrain the CPT for Tes_3 such that the explaining away effect can occur between Lie and Murder for value $\mathrm{Tes}_3 = true$. This is achieved by constraining the CPT for Tes_3 such that $\mathbf{X}^{-}(\{\mathrm{Lie}, \mathrm{Murder}\}, tes_3)$, as captured by step 6a of our approach.



		Murder
Mot_1	Mot_2	t
tt		0.8
tf		0.2
ft		0.2
ff		0.1

(b)

Fig. 12. BN graph constructed from the IG of Fig. 8b by our approach (a); a possible CPT for node Murder (b).



		Tes$_3$
Murder	Lie	t
ft		0.2
ff		0.8
tt		0.3
tf		0.01

(c)

Fig. 13. IG involving exceptions to generalisation arcs in $\mathbf{G^e}$ and $\mathbf{G^c}$ (a); the corresponding BN graph constructed by our approach (b); a possible CPT for node Tes$_3$ (c).

In this particular example, $\neg murder$ is one of the possible causes of tes_3; therefore, for variable Murder the ordering $false > true$ is assumed. For example, the CPT for Tes$_3$ can be chosen as in Fig. 13c, as in this case it holds that $\Pr(tes_3 \mid \neg murder, lie) \cdot \Pr(tes_3 \mid murder, \neg lie) = 0.2 \cdot 0.01 \leq \Pr(tes_3 \mid \neg murder, \neg lie) \cdot \Pr(tes_3 \mid murder, lie) = 0.8 \cdot 0.3$.

We note that multiple exceptions to an evidential generalisation arc g express different competing alternative explanations for the common effects expressed by $\mathbf{Tails}(g)$. We therefore propose to constrain the CPTs for the nodes corresponding to the tails such that a negative product synergy is exhibited between the nodes corresponding to each pair of exceptions, as captured by step 6a of our approach.

Step 3b In the IG of Fig. 13a, proposition $\neg opp$, which states that Marjan did not have opportunity to commit the murder as she has an alibi ($alibi$), provides an exception to the causal generalisation arc $mot_1 \rightarrow murder$. In contrast with the exception to the evidential generalisation arc, this exception cannot be considered a competing alternative explanation for the tail of the generalisation arc; the absence of opportunity cannot be considered a cause for motive. Instead, it allows us to infer that Marjan did not murder Leo ($\neg murder$). For exceptions to generalisations $g \in \mathbf{G^c}$, we therefore propose to form a head-to-head node in $\mathrm{Var}(Head(g))$ as opposed to in $\mathrm{Var}(p_i)$ for $p_i \in \mathbf{Tails}(g)$. By step 2b of our approach, the BN graph under construction includes arc Mot$_1$ $\rightarrow$ Murder. A head-to-head node can, therefore, be formed in Murder by adding additional arc Opp $\rightarrow$ Murder to the BN graph; this is captured by step 3b of our approach. The corresponding BN graph is depicted in Fig. 13b. As Murder describes both values $murder$ and $\neg murder$, possible interactions, if any, between mot_1 and $\neg opp$, and hence between Mot$_1$ and Opp, can be captured in the CPT for this node.

Steps 6b-c Bex and Renooij [5] previously noted that, for deduction, the presence of a proposition opposing an inference step from $q_1, \ldots, q_n$ to q should decrease the probability that q is true. We propose to take a similar approach for exceptions to causal generalisations used in performing inference. For deduction with a generalisation $q_1, \ldots, q_n \rightarrow q$ in $\mathbf{G^c}$ in presence of an exception p, we propose to constrain the CPT for $\mathrm{Var}(q)$ such that $\Pr(q \mid p, q_1, \ldots, q_n) < \Pr(q \mid \neg p, q_1, \ldots, q_n)$, as captured by step 6b of our approach. For abduction with a generalisation $q_1, \ldots, q_n \rightarrow q$ in $\mathbf{G^c}$, the probability that q_i is true given q should decrease in the presence of an exception p for $i = 1, \ldots, n$. Accordingly, we propose to constrain the probabilities of the BN such that $\Pr(q_i \mid p, q) < \Pr(q_i \mid \neg p, q)$, $i = 1, \ldots, n$, as captured by step 6c of our approach. The latter constraints cannot be directly imposed on the CPTs for nodes $\mathrm{Var}(p)$, $\mathrm{Var}(q)$, or $\mathrm{Var}(q_i)$, as nodes $\mathrm{Var}(q_i)$ and $\mathrm{Var}(p)$ are parents of node $\mathrm{Var}(q)$ by steps 2b and 3b of our approach. We note that approaches have been proposed that allow one to use this set of probability constraints in an elicitation procedure for obtaining the required local probability distributions [14].

7. Properties of the approach

In this section, we prove a number of formal properties of our approach. In Sect. 7.1, we study conditions on IGs under which the fully automatically constructed initial BN graph is guaranteed to be acyclic. In Sect. 7.2, we prove that, as intended, BN graphs constructed by our approach capture reasoning patterns similar to those that can be read from an IG given the evidence. In Sect. 7.3, we look into the size of the CPTs and complexity of probabilistic inference in BN graphs constructed by our approach. Finally, in Sect. 7.4, we look into mapping properties of our approach; specifically, we investigate conditions under which the same BN graph is constructed from different IGs by our approach, and discuss ways by which a distinction can be made in the (conditional) probabilities of the BN under construction.

7.1. Constructing acyclic graphs

In this section, we study conditions under which the initial graph constructed by steps 1–3 of our approach is guaranteed to be a DAG. Hence, under these conditions the (manual) verification step of whether the obtained graph contains cycles (part of step 4 of our approach) can be skipped.

Conditions a) and b) of Proposition 1 concern the existence of exception arcs in IGs. Specifically, cycles are possibly introduced within weakly connected components of the BN graph under construction in step 3 of our approach in case exception arcs exist within weakly connected components of IGs (condition a). Furthermore, cycles are also possibly introduced in the

BN graph from a node V_1 in one weakly connected component via a node V_2 in another weakly connected component in this step in case exception arcs exist between propositions in separate weakly connected components of IGs (condition b). Examples of IGs violating these conditions are provided after the formal result.

Proposition 1. *Consider IG $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$, and let $G^*_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}^*_{\mathcal{I}})$ be the possibly disconnected sub-graph of $G_{\mathcal{I}}$ with $\mathbf{A}^*_{\mathcal{I}} = \mathbf{A}_{\mathcal{I}} \setminus \mathbf{Exc}$. Let $\mathbf{C} = \{C = (\mathbf{P}^{\mathsf{C}}, \mathbf{A}^{\mathsf{C}}_{\mathcal{I}}) \mid \mathbf{P}^{\mathsf{C}} \subseteq \mathbf{P}, \mathbf{A}^{\mathsf{C}}_{\mathcal{I}} \subseteq \mathbf{A}^*_{\mathcal{I}}, C$ is a weakly connected component of $G^*_{\mathcal{I}}\}$ be the set of* IG components*. Assume that the following conditions are satisfied:*

a) For any IG component $C \in \mathbf{C}$, there does not exist an $exc\colon p \rightsquigarrow g$ in $\mathbf{Exc}$ with $p \in \mathbf{P}^{\mathsf{C}}$, $g \in \mathbf{A}^{\mathsf{C}}_{\mathcal{I}}$.
b) For every pair of IG components $C_1, C_2 \in \mathbf{C}$, there does not exist both an $exc_1\colon p_1 \rightsquigarrow g_1$ in $\mathbf{Exc}$ with $p_1 \in \mathbf{P}^{\mathsf{C}_1}$, $g_1 \in \mathbf{A}^{\mathsf{C}_2}_{\mathcal{I}}$ and an $exc_2\colon p_2 \rightsquigarrow g_2$ in $\mathbf{Exc}$ with $p_2 \in \mathbf{P}^{\mathsf{C}_2}$, $g_2 \in \mathbf{A}^{\mathsf{C}_1}_{\mathcal{I}}$.

Let $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ be the graph constructed from $G_{\mathcal{I}}$ according to steps 1–3 of our approach. Then $G_{\mathcal{B}}$ is a DAG.

Proof. By setting arcs in $\mathbf{A}_{\mathcal{B}}$ per step 2 of our approach, no cycles are introduced. Specifically, our non-repetitiveness and consistency assumptions (see Sect. 4.1) jointly assume that for every $p \in \mathbf{P}$ there does not exist a generalisation chain $[g_1, \ldots, g_m]$ with $p \in \mathbf{Tails}(g_1)$ such that either $Head(g_m) = p$ or $Head(g_m) = -p$. Therefore, no chain of arcs exists in $\mathbf{A}_{\mathcal{B}}$ from a node P to itself. The only other case in which cycles are possibly introduced in $G_{\mathcal{B}}$ is when a causal cycle exists in $G_{\mathcal{I}}$, which is also prohibited by assumption (see Sect. 4.1).

We now prove that if $C \in \mathbf{C}$ is an IG component of $G_{\mathcal{I}}$, then the BN segment C' obtained from C after step 2 is a weakly connected component of the thus far constructed BN graph $G_{\mathcal{B}}$. Let $C \in \mathbf{C}$ be an IG component of $G_{\mathcal{I}}$. Then propositions within C are interconnected by arcs in $\mathbf{G}$ and $\mathbf{N}$ but are not connected to other propositions in the supergraph $G_{\mathcal{I}}$; therefore, corresponding nodes in BN segment C' are interconnected but not connected to other nodes in supergraph $G_{\mathcal{B}}$. This is the case as per step 2, $\mathbf{A}_{\mathcal{B}}$ only includes arcs between the variables corresponding to $\mathbf{Tails}(g)$ and $Head(g)$ for every $g \in \mathbf{G}$; no arcs are introduced corresponding to $n \in \mathbf{N}$. We then call C' the weakly connected component *corresponding to* IG component C. In step 3 of our approach, additional arcs are included in $\mathbf{A}_{\mathcal{B}}$ for every $exc \in \mathbf{Exc}$. We now prove that no cycles are introduced within the weakly connected components of $G_{\mathcal{B}}$ or from a node V_1 in one weakly connected component to itself via a node V_2 in another weakly connected component of $G_{\mathcal{B}}$ in step 3. Under condition a), no cycles are introduced within a weakly connected component C' of $G_{\mathcal{B}}$ in this step. Specifically, C' contains no cycles after step 2 and no cycles are introduced in C' in step 3 as no exception arc is directed from a $p \in \mathbf{P}^{\mathsf{C}}$ to a $g \in \mathbf{A}^{\mathsf{C}}_{\mathcal{I}}$ in corresponding IG-component C. Furthermore, for every pair of IG components C_1 and C_2 of $G_{\mathcal{I}}$ with corresponding weakly connected components C'_1 and C'_2 of $G_{\mathcal{B}}$, no cycles are introduced from a node $V_1 \in C'_1$ to itself via a node $V_2 \in C'_2$ under condition b). The resulting BN graph is therefore acyclic. □

Figs. 14a, 14c and 14e depict examples of IGs that do not satisfy condition a) of Proposition 1 and hence result in cyclic graphs. In general, an IG violating only condition a) either contains:

(1a) A generalisation chain $[g_1, \ldots, g_m]$, $g_1, \ldots, g_m \in \mathbf{G}^{\mathsf{c}}$ and an exception arc $exc\colon Head(g_j) \rightsquigarrow g_i$ for $1 \le i < j \le m$ (see Figs. 14a and 14c), or;
(1b) A generalisation chain $[g_1, \ldots, g_m]$, $g_1, \ldots, g_m \in \mathbf{G}^{\mathsf{e}}$ and an exception arc $exc\colon Head(g_i) \rightsquigarrow g_j$ for $1 \le i < j \le m$, or;
(2) Propositions r, $\neg r$ with $n\colon r \leftrightsquigarrow \neg r$ in $\mathbf{N}$, where $\neg r$ provides an exception to a generalisation g_i in a generalisation chain $[g_1, \ldots, g_m]$ with either:
  (2a) $Head(g_m) = r$ and $g_1, \ldots, g_m \in \mathbf{G}^{\mathsf{c}}$ (see Fig. 14e), or;
  (2b) $r \in \mathbf{Tails}(g_1)$ and $g_1, \ldots, g_m \in \mathbf{G}^{\mathsf{e}}$.

For 1a), $Head(g_j)$ poses an exception to a generalisation that was used in iteratively inferring $Head(g_j)$ in case solely deductive inferences are performed with the generalisations in the chain, as illustrated in Fig. 14a. However, in case abductive inferences are performed with generalisations in the chain, it may not be the case that $Head(g_j)$ poses an exception to a generalisation that was used in iteratively inferring $Head(g_j)$, as illustrated in Fig. 14c. For 1b) $Head(g_i)$ poses an exception to a generalisation that is used to iteratively deductively infer another proposition from $Head(g_i)$, as only deduction can be performed with evidential generalisations. For (2), the question remains whether realistic examples of IGs including such conflict relations can be constructed; an abstract example is provided in Fig. 14e. Condition a) of Proposition 1 thus mostly poses a technical constraint to ensure acyclic graphs are constructed by our approach.

IGs violating condition b) may appear more frequently; an example is provided in Fig. 14g. In the validation step that follows the initial construction of BN graphs corresponding to IGs violating conditions a) and b), arcs can be reversed or removed to make these graphs acyclic. The choice of arc to reverse or remove will depend on its effect on active chains, including those between nodes not directly incident on the arc. We note that this type of (manual) verification is standard in BN construction, especially in data-poor domains (see Sect. 5.3). While the domain knowledge expressed in the original

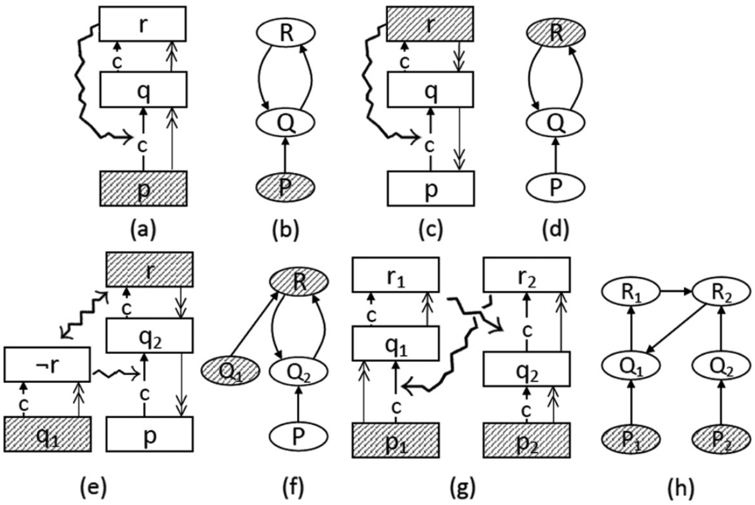


Fig. 14. Examples of IGs (a, c, e, g) for which a cyclic graph is constructed by steps 1–3 of our approach (b, d, f, h).

IG has been exploited to construct an initial BN graph, additional domain knowledge may need to be elicited to obtain a valid graph.

7.2. Capturing induced reasoning patterns expressed by IGs as active chains

In this section, we study whether BN graphs constructed by our approach capture reasoning patterns similar to those that can be read from the original IG given the evidence. As motivated throughout this paper, generalisation arcs only capture knowledge about the world in conditional form; only when considering the available evidence $\mathbf{E_p}$ in the IG can directionality of inference be read from the graph. Specifically, in Definitions 9 and 10 conditions are specified under which (a set) of proposition(s) are deductively respectively abductively inferred from another (set of) proposition(s) given $\mathbf{E_p}$. The following notion of an *inference chain* describes a sequence of propositions that are iteratively inferred from each other given $\mathbf{E_p}$. For those familiar with argumentation, we note that inference chains are comparable to arguments as defined in ASPIC$^+$ [30]. A key distinction is that we define inference *chains* and not inference *trees* (which would more closely resemble arguments), as our current focus lies on defining a concept that is more closely related to the concept of active chains for BNs. Furthermore, the notion of attack is not required for our current purposes. In previous work [42], we investigated the relations between argumentation and inference as it can be performed with our IG-formalism; more specifically, it is shown that an Argumentation Framework (AF) as in Dung [15] can straightforwardly be generated from an IG by considering the available evidence. For details, the reader is referred to [42].

First, we define the concept of a *chain* for IGs.

Definition 21 *(Chain in an IG).* Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $\{p_1, \dots, p_n\} \subseteq \mathbf{P}$ and let $\mathbf{G}' = \{g_1, \dots, g_{n-1}\} \subseteq \mathbf{G}$. Then $(p_1, g_1, p_2, g_2, \dots, p_{n-1}, g_{n-1}, p_n)$ is a *chain in* $G_{\mathcal{I}}$ iff for all $1 < i \leq n$ it either holds that $Head(g_{i-1}) = p_i,\ p_{i-1} \in \mathbf{Tails}(g_{i-1})$ or $Head(g_{i-1}) = p_{i-1},\ p_i \in \mathbf{Tails}(g_{i-1})$.

We now define when a chain in an IG is an inference chain.

Definition 22 *(Inference chain).* Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $\mathbf{G}' = \{g_1, \dots, g_{n-1}\} \subseteq \mathbf{G}$, and let $\{p_1, \dots, p_n\} \subseteq \mathbf{P}$ such that $\nexists i, j \in \{1, \dots, n\}$ with $p_i = -p_j$, and such that $(p_1, g_1, p_2, g_2, \dots, p_{n-1}, g_{n-1}, p_n)$ is a chain in $G_{\mathcal{I}}$. Let $p_1 \in \mathbf{E_p}$ or let p_1 be deductively or abductively inferred using a generalisation $g \in \mathbf{G} \setminus \mathbf{G}'$ given $\mathbf{E_p}$ (see Definitions 9 and 10). Then chain $(p_1, g_1, p_2, g_2, \dots, p_{n-1}, g_{n-1}, p_n)$ is an *inference chain* in $G_{\mathcal{I}}$ given $\mathbf{E_p}$ iff for all $1 < i \leq n$ it holds that:

1. p_i is deductively inferred using generalisation $g_{i-1} \in \mathbf{G}'$ given $\mathbf{E_p}$ (see Definition 9), where $Head(g_{i-1}) = p_i,\ p_{i-1} \in \mathbf{Tails}(g_{i-1})$, or;
2. p_i is abductively inferred from p_{i-1} using generalisation $g_{i-1} \in \mathbf{G}'$ given $\mathbf{E_p}$ (see Definition 10), where $Head(g_{i-1}) = p_{i-1}$, $p_i \in \mathbf{Tails}(g_{i-1})$.

We emphasise that an inference chain $(p_1, g_1, p_2, g_2, \dots, p_{n-1}, g_{n-1}, p_n)$ does not only describe that p_{i-1} was used in inferring p_i for all $1 < i \leq n$; it also describes that the inference chain needs to start in a proposition p_1 that is either observed or inferred, hence the conditions regarding p_1 in Definition 22. We refer to the assumption that for inference chains $(p_1, g_1, p_2, g_2, \dots, p_{n-1}, g_{n-1}, p_n)$ it holds that all p_i are distinct (enforced by assuming that $\{p_1, \dots, p_n\} \subseteq \mathbf{P}$) as

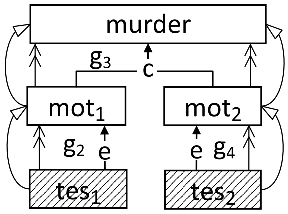


Fig. 15. The IG of Fig. 8b, where inference chains are also indicated by connecting arcs with open arrowheads.

our *non-repetitiveness* assumption on inference chains. We refer to the assumption that for $\{p_1, \ldots, p_n\}$ it holds that $\nexists i, j \in \{1, \ldots, n\}$ with $p_i = -p_j$ as our *consistency* assumption on inference chains.

Compared to generalisation chains (see Definition 4), which are solely captured by the graphical structure of IGs, inference chains can only be read from an IG by considering the evidence $\mathbf{E_p}$. In case an inference chain only describes deductive inference steps, then our non-repetitiveness and consistency assumptions on inference chains coincide with our non-repetitiveness and consistency assumptions on generalisation chains as described in Sect. 4.1; however, these assumptions do not coincide in case an inference chain also describes abductive inference steps.

The following example illustrates the concept ‘inference chain’ and how it compares to the concept ‘generalisation chain’.

Example 22. In the IG of Fig. 15, (tes_1, g_2, mot_1, g_3, $murder$) is an inference chain given $\mathbf{E_p}$, as mot_1 is deductively inferred from $tes_1 \in \mathbf{E_p}$ using g_2, where $Head(g_2) = mot_1$ and $tes_1 \in \mathbf{Tails}(g_2)$, and $murder$ is deductively inferred from mot_1 and mot_2 using g_3, where $Head(g_3) = murder$ and $mot_1 \in \mathbf{Tails}(g_3)$. In this IG, $[g_2, g_3]$ is also a generalisation chain (see Example 8). Note that the presence of this inference chain does not imply that mot_1 is by itself sufficient to infer $murder$; instead, $murder$ can only be deductively inferred using g_3 in case both mot_1 *and* mot_2 are affirmed. The broader context in which the inference step from mot_1 to $murder$ is performed using g_3 is thus not directly apparent from this inference chain; instead, the role of proposition mot_2 becomes apparent in considering other inference chains that can be read from this IG given $\mathbf{E_p}$, specifically inference chain (tes_2, g_4, mot_2, g_3, $murder$).

In the IG of Fig. 4, ($police$, g_1, $murder$, g_3, mot_1) is an inference chain given $\mathbf{E_p}$: $murder$ is deductively inferred from $police \in \mathbf{E_p}$ using generalisation g_1 and mot_1 is abductively inferred from $murder$ using generalisation g_3. However, $[g_1, g_3]$ is not a generalisation chain, as $Head(g_1) = murder \notin \mathbf{Tails}(g_3)$. □

To remain closely related to the concept of active chains for BNs, we assume in Definition 22 that inference chains $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ do not need to start in evidence in that it does not need to hold that $p_1 \in \mathbf{E_p}$, as long as p_1 is deductively or abductively inferred using a $g \in \mathbf{G} \setminus \mathbf{G}'$ given $\mathbf{E_p}$.

Example 23. In Fig. 15, (mot_1, g_3, $murder$) is an inference chain given $\mathbf{E_p}$: $murder$ is deductively inferred from mot_1 and mot_2 using g_3. However, $mot_1 \notin \mathbf{E_p}$; instead, mot_1 is deductively inferred using $g_2 \in \mathbf{G} \setminus \{g_3\}$ given $\mathbf{E_p}$. □

The following example illustrates that inference chains are generally not symmetrical, in contrast with active chains for BNs.

Example 24. In the IG of Fig. 15, (tes_1, g_2, mot_1, g_3, $murder$) is an inference chain (see Example 22), but ($murder$, g_3, mot_1, g_2, tes_1) is not an inference chain as mot_1 cannot be inferred from $murder$ using g_3 and tes_1 cannot be inferred from mot_1 using g_2. □

We prove the following properties of inference chains. Lemma 1 states that for inference chains, only the first proposition in the chain can possibly be observed.

Lemma 1. *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $p_1, \ldots, p_n \in \mathbf{P}$, $g_1, \ldots, g_{n-1} \in \mathbf{G}$ and let $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ be an inference chain in $G_{\mathcal{I}}$ given $\mathbf{E_p}$. Then $p_i \notin \mathbf{E_p}$ for $i > 1$.*

Proof. Let $i > 1$. In case p_i is deductively inferred from p_{i-1} using g_{i-1}, then $p_i = Head(g_{i-1}) \notin \mathbf{E_p}$ per the restrictions of Definition 9. Similarly, in case p_i is abductively inferred from p_{i-1} using g_{i-1}, then $p_i \notin \mathbf{E_p}$, as $p_i \in \mathbf{Tails}(g_{i-1})$ and $\mathbf{Tails}(g_{i-1}) \cap \mathbf{E_p} = \emptyset$ per the restrictions of Definition 10. □

Lemma 2 states that an inference step between two consecutive propositions p_i and p_{i+1} in an inference chain can only be performed with a generalisation g_i for which $Head(g_i) = p_i$ and $p_{i+1} \in \mathbf{Tails}(g_i)$ in case g_i is a *causal* generalisation.

Lemma 2. *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $p_1, \ldots, p_n \in \mathbf{P}$, $g_1, \ldots, g_{n-1} \in \mathbf{G}$ and let $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ be an inference chain in $G_{\mathcal{I}}$ given $\mathbf{E_p}$. Let $i \in \{1, \ldots, n-1\}$ and assume that $Head(g_i) = p_i$, $p_{i+1} \in \mathbf{Tails}(g_i)$. Then $g_i \in \mathbf{G}^{\mathrm{c}}$.*

Proof. Assume a generalisation g_i with $Head(g_i) = p_i$ and $p_{i+1} \in \mathbf{Tails}(g_i)$ is indicated in $G_{\mathcal{I}}$, then p_{i+1} cannot be inferred from p_i in case $g_i \in \mathbf{G}^{\mathrm{e}}$, as this would be an instance of abductive inference while per the restrictions of Definition 10 abduction can only be performed using generalisation arcs in $\mathbf{G}^{\mathrm{c}}$. □

In performing inference care should be taken that no cause for an effect is inferred if an alternative cause for this effect was already previously inferred (Pearl's constraint, see Sect. 2.4). In the context of IGs, for $g \in \mathbf{G}^{\mathrm{c}}$, propositions in $\mathbf{Tails}(g)$ express a cause for the common effect expressed by $Head(g)$, and for $g \in \mathbf{G}^{\mathrm{e}}$, $Head(g)$ expresses the usual cause for propositions in $\mathbf{Tails}(g)$. Hence, in defining how inferences can be read from IGs, restrictions are put in Definitions 9 and 10 such that Pearl's constraint is adhered to. We now formally prove that the inference chains that can be read from an IG given an evidence set $\mathbf{E_p}$ indeed never violate Pearl's constraint.

First, we formally define Pearl's constraint in the context of IGs.

Definition 23 *(Pearl's constraint).* Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $c_1, c_2 \in \mathbf{P}$ be alternative causes of $e \in \mathbf{P}$, as indicated by generalisations $g, g' \in \mathbf{G}$ (see Definition 3). Then chain (c_1, g, e, g', c_2) is not an inference chain in $G_{\mathcal{I}}$ given $\mathbf{E_p}$.

We now formally prove that Pearl's constraint is indeed adhered to.

Proposition 2 *(Adherence to Pearl's constraint). Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG, and let $\mathbf{E_p} \subseteq \mathbf{P}$ be an evidence set. Let $c_1, c_2 \in \mathbf{P}$ be alternative causes of $e \in \mathbf{P}$, as indicated by generalisations $g, g' \in \mathbf{G}$ (see Definition 3). Then Pearl's constraint is adhered to.*

Proof. We need to prove that chain (c_1, g, e, g', c_2) is not an inference chain in $G_{\mathcal{I}}$ given $\mathbf{E_p}$. In performing the inference step from c_1 to e, a generalisation $g \in \mathbf{G}^{\mathrm{e}}$, $Head(g) = c_1$, $e \in \mathbf{Tails}(g)$ could not have been used (case 1 of Definition 3) per Lemma 2. Thus, we only need to consider case 2 of Definition 3, which is a deductive inference step. First, consider case 2a of Definition 3. Then by Definition 10 (condition 2), c_2 cannot be inferred from e using g'. Next, consider case 2b of Definition 3. Then by Definition 9 (condition 2), c_2 cannot be inferred from e using g'. □

Example 25. In the IG of Fig. 5c, $[g_1, g_2]$ is a generalisation chain but (*smoke_machine*, g_1, *smoke*, g_2, *fire*) is *not* an inference chain, as per Pearl's constraint *fire* cannot be deductively inferred from *smoke* using g_2. □

An IG, by means of its inference chains, describes sequences of propositions that can be iteratively inferred from each other given the available evidence. In comparison, from a BN graph we can read the chains between nodes that are active given the evidence and will be exploited to propagate the evidence upon probabilistic inference. We now formally prove that all inference chains that can be read from an IG given the evidence are captured in the BN graph by means of active chains given the available evidence for $\mathbf{E_V} \subseteq \mathbf{V}$ corresponding to $\mathbf{E_p} \subseteq \mathbf{P}$. This result implies that, for every inference chain $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ given $\mathbf{E_p}$, nodes $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$ are not d-separated given the evidence for $\mathbf{E_V}$.

Proposition 3. *Let $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}})$ be an IG with evidence set $\mathbf{E_p} \subseteq \mathbf{P}$, and let $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ be the BN graph constructed from $G_{\mathcal{I}}$ according to steps 1–3 of our approach. Let $(p_1, g_1, p_2, g_2, \ldots, p_{n-1}, g_{n-1}, p_n)$ be any inference chain that can be read from $G_{\mathcal{I}}$ given $\mathbf{E_p}$. Then there exists an active chain between $Var(p_1)$ and $Var(p_n)$ in $G_{\mathcal{B}}$ given the evidence for $\mathbf{E_V}$.*

Proof. Following steps $1-2$ of our approach, a sequence of nodes and arcs is formed between $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$ in $G_{\mathcal{B}}$, as for every g_i, $1 \leq i < n$ arcs between $\mathbf{Tails}(g_i)$ and $Head(g_i)$ are added to $\mathbf{A}_{\mathcal{B}}$. By our non-repetitiveness and consistency assumptions on inference chains, this is a sequence of *distinct* nodes and arcs and thus a chain in $G_{\mathcal{B}}$. We now prove that this chain in the BN graph is active given $\mathbf{E_V}$, as all options to block a chain do not occur. First, note that per Lemma 1 it holds that $p_i \notin \mathbf{E_p}$ for $i > 1$; therefore, corresponding nodes $\mathrm{Var}(p_i)$ in the BN graph are not instantiated and hence do not block chains. Possibly only $p_1 \in \mathbf{E_p}$. However, in this case, the corresponding node $\mathrm{Var}(p_1)$ is an end-point of the chain which, therefore, does not block it. Hence, chains between $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$ are never blocked by $\mathbf{E_V}$.

The only other option to block a chain occurs in case it includes an uninstantiated head-to-head node without instantiated descendants. Consider p_{i-1}, p_i, p_{i+1} for an arbitrary $1 < i < n$, and let g_{i-1} and g_i be the corresponding generalisations used in the inferences from p_{i-1} to p_i and from p_i to p_{i+1}, respectively. We show that a head-to-head node $\mathrm{Var}(p_{i-1}) \rightarrow \mathrm{Var}(p_i) \leftarrow \mathrm{Var}(p_{i+1})$ is never formed. Note that by steps 2*a* and 2*b* of our approach, a head-to-head node $\mathrm{Var}(p_{i-1}) \rightarrow \mathrm{Var}(p_i) \leftarrow \mathrm{Var}(p_{i+1})$ is only formed in case:

1. $g_{i-1} \in \mathbf{G}^{\mathrm{e}}$, $Head(g_{i-1}) = p_{i-1}$, $p_i \in \mathbf{Tails}(g_{i-1})$, and either:
   1a) $g_i \in \mathbf{G}^{\mathrm{e}}$, $Head(g_i) = p_{i+1}$, $p_i \in \mathbf{Tails}(g_i)$, or;

1b) $g_i \in \mathbf{G}^\mathrm{c}$, $Head(g_i) = p_i$, $p_{i+1} \in \mathbf{Tails}(g_i)$.

2. $g_{i-1} \in \mathbf{G}^\mathrm{c}$, $Head(g_{i-1}) = p_i$, $p_{i-1} \in \mathbf{Tails}(g_{i-1})$, and either:
   2a) $g_i \in \mathbf{G}^\mathrm{c}$, $Head(g_i) = p_i$, $p_{i+1} \in \mathbf{Tails}(g_i)$, or;
   2b) $g_i \in \mathbf{G}^\mathrm{e}$, $Head(g_i) = p_{i+1}$, $p_i \in \mathbf{Tails}(g_i)$.

However, in performing the inference steps from p_{i-1} to p_i and from p_i to p_{i+1} none of these combinations of generalisations could have been used, as proven in Proposition 2. Thus a head-to-head node $\mathrm{Var}(p_{i-1}) \to \mathrm{Var}(p_i) \leftarrow \mathrm{Var}(p_{i+1})$ is never formed, and chains between $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$ are never blocked.

Finally, in step 3 $\mathbf{A}_\mathcal{B}$ is extended for exception arcs. This step does not change the chains formed between $\mathrm{Var}(p_1)$ and $\mathrm{Var}(p_n)$ in step 2, which therefore remain active given $\mathbf{E_V}$. $\square$

The implication in the other direction of Proposition 3 does not generally hold. Specifically, it does not generally hold that for every induced active chain in a BN graph constructed from an IG $G_\mathcal{I}$, there exists a corresponding induced inference chain in $G_\mathcal{I}$. For instance, since the notion of an active chain is a symmetrical concept, a BN graph will also capture reasoning patterns in the direction opposite of the inference chains that can be read from an IG. As inference chains are generally not symmetrical (see Example 24), reasoning patterns may appear in the BN graph that do not appear in the original IG.

7.3. Size and complexity of constructed BNs

The following properties concern the size and complexity of the resulting BN model. Proposition 4 gives an upper-bound on the total number of nodes and arcs introduced in a BN graph constructed from an IG by our approach.

Proposition 4. *Let $G_\mathcal{I} = (\mathbf{P}, \mathbf{A}_\mathcal{I})$ be an IG, and let $G_\mathcal{B} = (\mathbf{V}, \mathbf{A}_\mathcal{B})$ be the BN graph constructed from $G_\mathcal{I}$ according to steps 1–3 of our approach. Let $\mathbf{Exc}^\mathrm{e}$ and $\mathbf{Exc}^\mathrm{c}$ be disjoint subsets of $\mathbf{Exc}$ consisting of exceptions to generalisation arcs in $\mathbf{G}^\mathrm{e}$ and $\mathbf{G}^\mathrm{c}$, respectively. Then $|\mathbf{V}| = |\mathbf{P}| - |\{p \mid p \in \mathbf{P} \text{ and } \neg p \in \mathbf{P}\}|$ and $|\mathbf{A}_\mathcal{B}| \le \sum_{g \in \mathbf{G}} |\mathbf{Tails}(g)| + |\mathbf{Exc}^\mathrm{c}| + \sum_{p \rightsquigarrow g \text{ in } \mathbf{Exc}^\mathrm{e}} |\mathbf{Tails}(g)|$.* $\square$

Proof. By step 1 of our approach, both p and its negation are mapped to the same node $\mathrm{Var}(p) = \mathrm{Var}(\neg p) \in \mathbf{V}$. Therefore, the exact number of nodes introduced in this step is $|\mathbf{P} \setminus \{p \mid p \in \mathbf{P} \text{ and } -p \in \mathbf{P}\}|$. In step 2, at most $|\mathbf{Tails}(g)|$ arcs are added to $\mathbf{A}_\mathcal{B}$ for every $g \in \mathbf{G}$. For every $exc \in \mathbf{Exc}^\mathrm{c}$, one additional arc is added to $\mathbf{A}_\mathcal{B}$ in step 3b. For every $exc\colon p \rightsquigarrow g$ in $\mathbf{Exc}^\mathrm{e}$, at most $|\mathbf{Tails}(g)|$ arcs are added to $\mathbf{A}_\mathcal{B}$ in step 3a. $\square$

As a corollary, note that the complexity of constructing a BN graph from an IG using our approach is linear in the number of proposition nodes, generalisation arcs and exceptions arcs in the IG, as nodes in the BN graph are directly added according to the IG's proposition nodes and arcs in the BN graph are directly added according to the IG's generalisation arcs and exception arcs.

Proposition 5 gives an upper-bound on the number of parents introduced by our approach for each node $\mathrm{Var}(p)$ in $\mathbf{V}$, which bounds both the size of the CPTs and the complexity of probabilistic inference in the BN [11, pp. 141–142]. Informally, this bound captures the number of generalisation arcs and exception arcs that involve either proposition p or $\neg p$. The terminology used in Proposition 5 is illustrated in Fig. 16.

Proposition 5. *Let $G_\mathcal{I} = (\mathbf{P}, \mathbf{A}_\mathcal{I})$ be an IG, and let $G_\mathcal{B} = (\mathbf{V}, \mathbf{A}_\mathcal{B})$ be the BN graph constructed from $G_\mathcal{I}$ according to steps 1–3 of our approach. For every $p \in \mathbf{P}$, let $\mathbf{Par}_p = \{p_i \mid p_i \in \mathbf{Tails}(g), g \in \mathbf{G}^\mathrm{c}, Head(g) \in \{p, \neg p\}\}$. Let $\mathbf{G}^\mathrm{e}_p$ be a subset of $\mathbf{G}^\mathrm{e}$, where $g \in \mathbf{G}^\mathrm{e}_p$ iff $p \in \mathbf{Tails}(g)$. Let $\mathbf{Exc}_p \subseteq \mathbf{Exc}$ be the subset of exception arcs directed to a $g \in \mathbf{G}^\mathrm{e}_p$ or a $g \in \mathbf{G}^\mathrm{e}_{\neg p}$. Similarly, let $\mathbf{Exc}'_p \subseteq \mathbf{Exc}$ be the subset of exception arcs directed to a $g \in \mathbf{G}^\mathrm{c}$ for which $Head(g) \in \{p, \neg p\}$. Then an upper-bound for the number of parents of $\mathrm{Var}(p)$ is:*

$$|\mathbf{Par}_p| + |\mathbf{Exc}_p| + |\mathbf{Exc}'_p| + |\mathbf{G}^\mathrm{e}_p| + |\mathbf{G}^\mathrm{e}_{\neg p}|$$

Proof. For every $g \in \mathbf{G}^\mathrm{c}$ with $Head(g) \in \{p, \neg p\}$, $\mathrm{Var}(p)$ has at most $|\mathbf{Tails}(g)|$ parents by step 2b of our approach; hence the term $|\mathbf{Par}_p|$. By steps 3a and 3b, $\mathbf{A}_\mathcal{B}$ includes a single arc directed towards $\mathrm{Var}(p)$ for every exception exc in $\mathbf{Exc}_p$ or in $\mathbf{Exc}'_p$, respectively; hence the terms $|\mathbf{Exc}_p|$ and $|\mathbf{Exc}'_p|$. For every $g \in \mathbf{G}^\mathrm{e}$ with p or $\neg p$ in $\mathbf{Tails}(g)$, a single arc directed towards $\mathrm{Var}(p)$ is included in $\mathbf{A}_\mathcal{B}$ by step 2a of our approach; hence the terms $|\mathbf{G}^\mathrm{e}_p|$ and $|\mathbf{G}^\mathrm{e}_{\neg p}|$. $\square$

Note that, in case $\mathbf{G}^\mathrm{e}_p = \mathbf{G}^\mathrm{e}_{\neg p} = \emptyset$, it follows that $\mathbf{Exc}_p = \emptyset$; hence, terms $|\mathbf{G}^\mathrm{e}_p|$, $|\mathbf{G}^\mathrm{e}_{\neg p}|$ and $|\mathbf{Exc}_p|$ are equal to zero in this case. Similarly, $\mathbf{Par}_p$ may be empty, in which case $\mathbf{Exc}'_p = \emptyset$ and terms $|\mathbf{Par}_p|$ and $|\mathbf{Exc}'_p|$ are equal to zero.

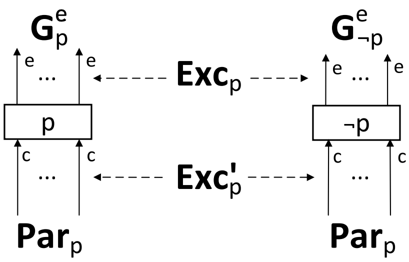


Fig. 16. Illustration of the terminology used in Proposition 5.

7.4. Mapping properties and probabilistic constraints

Finally, we investigate conditions under which the same BN graph is constructed from different IGs by our approach, and discuss ways by which a distinction can be made between these different cases in the (conditional) probabilities of the BN. First, we prove in Proposition 6 that for every finite BN graph $G_{\mathcal{B}}$, there exists a finite IG such that this IG is mapped to $G_{\mathcal{B}}$ by our approach.

Proposition 6. *Let* **IG** *be the space of finite IGs and let* **BN** *be the space of finite BN graphs. Let $\mathcal{F}$:* **IG** $\rightarrow$ **BN** *be the function defined by steps 1–3 of our approach. Then $\mathcal{F}$ is a surjection.*

Proof. Let $G_{\mathcal{B}} = (\mathbf{V}, \mathbf{A}_{\mathcal{B}})$ be a BN graph in **BN**. Then we need to find at least one IG $G_{\mathcal{I}} = (\mathbf{P}, \mathbf{A}_{\mathcal{I}}) \in \mathbf{IG}$ s.t. $\mathcal{F}(G_{\mathcal{I}}) = G_{\mathcal{B}}$. Define $G_{\mathcal{I}}$ as follows. For every node $\mathrm{P} \in \mathbf{V}$, include proposition $p \in \mathbf{P}$. For every arc $\mathrm{P}_1 \rightarrow \mathrm{P}_2 \in \mathbf{A}_{\mathcal{B}}$, include generalisation arc $g\colon p_1 \rightarrow p_2$ in $\mathbf{G}^{\mathrm{c}}$. Then $\mathcal{F}(G_{\mathcal{I}}) = G_{\mathcal{B}}$ by steps 1 and 2b. □

However, $\mathcal{F}$ is not an injection. Figs. 17a-d depict examples of IGs for which the same BN graph, namely the graph depicted in Fig. 17e, is constructed by $\mathcal{F}$. Possible differences between these IGs can be captured in the (conditional) probabilities of the BN under construction. In Fig. 17a, a negation arc is drawn between r and $\neg r$. A possible probabilistic interpretation is that this IG informs us on probabilities $\Pr(r \mid p, q)$ and $\Pr(\neg r \mid p, q)$, where a preference for r over $\neg r$ defines an ordering on these two probabilities. In our IG-formalism, we opted not to account for preferences, as these are typically not indicated using reasoning tools; hence, possible probabilistic constraints resulting from such preferences are not further discussed.

In Fig. 17b, p and q can each be considered sufficient for deductively inferring r, while in Fig. 17c both p and q are needed. A possible probabilistic interpretation of the IG in Fig. 17b is that it only informs us on probabilities $\Pr(r \mid p)$ and $\Pr(r \mid q)$ and not on $\Pr(r \mid p, q)$, while the reverse holds for the IG of Fig. 17c. Fig. 17c is distinguished from Fig. 17a, as Fig. 17c only informs us on $\Pr(r \mid p, q)$ while Fig. 17a also informs us on $\Pr(r \mid q)$ and $\Pr(\neg r \mid p)$. For exception arcs, specific probabilistic constraints are derived, as captured by step 6 of our approach. Specifically, in the example of Fig. 17d, constraint $\Pr(r \mid p, q) < \Pr(r \mid \neg p, q)$ is derived. Related research on the relations between probability and inference is discussed in Sect. 9.3.

8. Case study: the Sacco and Vanzetti case

In this section, we apply our BN graph construction approach to parts of an actual legal case, namely the well-known Sacco and Vanzetti case. The case concerns Sacco and Vanzetti, who were convicted for shooting and killing payroll guard Berardelli during a robbery in South Braintree, Massachusetts on 15 April 1920; a detailed description of the case is provided by Kadane and Schum [22]. Kadane and Schum performed a probabilistic analysis of this case by first constructing

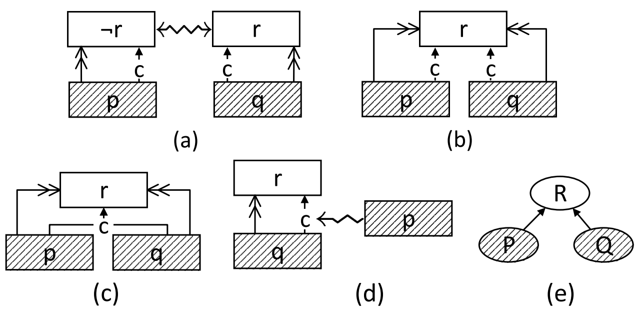


Fig. 17. Examples of IGs (a-d) for which the same BN graph (e) is constructed by our approach.

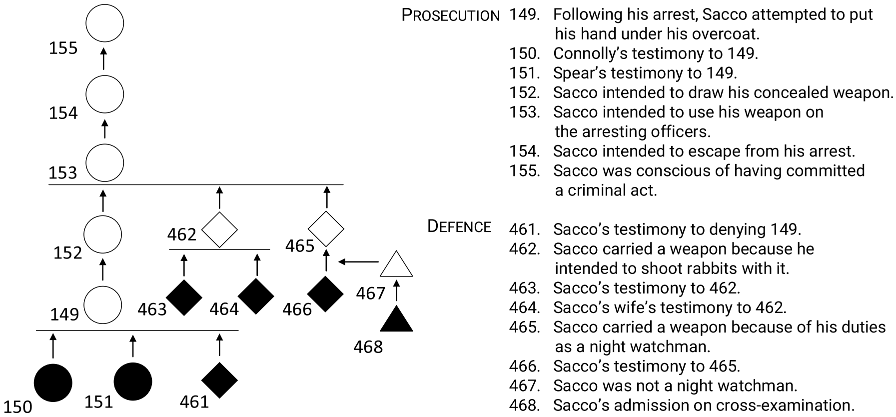


Fig. 18. Wigmore chart concerning Sacco's consciousness of guilt, along with the corresponding key list, adapted from Kadane and Schum [22, pp. 330–331].

Wigmore charts [43] (described below) of aspects of the case and then manually constructed corresponding BNs by assessing the modelled independence relation and assessing the necessary (conditional) probabilities. In this section, we illustrate and perform a first validation of our approach by formalising one of Kadane and Schum's Wigmore charts (chart 25, [22, pp. 330–331]) as an IG, where we compare the obtained BN graph to their BN graph. The currently presented case study is an extension of the case study that appeared in our previous work [39] in which parts of the case were interpreted as a preliminary version of an IG in which the roles of generalisation and inference are not separated. In the current paper, we describe our IG modelling choices in more detail and we provide a more detailed comparison of the BN graph constructed by our approach to that of [22].

This section is structured as follows. In Sect. 8.1 Kadane and Schum's Wigmore chart concerning Sacco's consciousness of guilt is presented, where a possible formalisation of this Wigmore chart as an IG is provided in Sect. 8.2. In Sect. 8.3 we then apply our BN graph construction approach to this IG and compare the obtained BN graph to that of Kadane and Schum. In Sect. 8.4 we then conclude the case study.

8.1. Wigmore chart concerning Sacco's consciousness of guilt

According to Kadane and Schum, the ultimate claim under consideration in the Sacco and Vanzetti case is Π_3, which states that *'It was Sacco who, with the assistence of Vanzetti, intentionally fired shots that took the life of Berardelli during the robbery and shooting that took place in South Brain tree.'* In the prosecution's case against Sacco and Vanzetti, their alleged consciousness of guilt in the South Braintree crime played an important role. However, as noted by Kadane and Schum the inferences made based on the available evidence for this part of the case are not particularly strong; a significant part of Kadane and Schum's analysis is, therefore, devoted to this part of the case. During their arrest, Sacco and Vanzetti were armed. According to the two arresting officers, Connolly and Spear, Sacco and Vanzetti made suspicious hand movements, from which the prosecution concluded that they intended to draw their concealed weapons in order to escape their arrest. This suggests that they were conscious of having committed a criminal act. In the remainder of this section, we only consider this part of the case.

In Fig. 18, a modernised Wigmore chart concerning Sacco's consciousness of guilt is depicted, adapted from Kadane and Schum [22, pp. 330–331]. Wigmore charts are diagrams familiar to many legal experts in which symbols indicating hypotheses and claims are joined by lines representing relations between these hypotheses and claims. Wigmore charts were introduced by John Henry Wigmore [43] and were further developed and studied from an academic perspective by the so-called 'New Evidence Theorists' including Anderson, Schum and Twining (see [22, pp. 70–71]), who provided a modernised, more user-friendly version of Wigmore's charting method. Wigmore introduced his method as an aid in structuring a mass of evidence in a legal case in detailed way. An important aspect of his method is that it not only used for expressing supporting reasons but also for revealing possible sources of doubt. Wigmore's charts can be considered a precursor of diagrams in argument diagramming tools [26], as well as a forerunner of instantiations of formal argumentation systems [1].

Compared to Kadane and Schum's original chart, we consider a subset of the mapped claims; in particular, claims 469, 470, 155a, 156 and Π_3, additional claims regarding Sacco's political beliefs (claims 471 – 480 in the original chart), and claims that were provided post-trial by historians are not considered. On the right-hand side of Fig. 18 the corresponding key list is depicted, which indicates for every number in the chart to which claim it corresponds. As noted by Kadane and Schum [22, p. 88], vertical arcs between nodes in their version of Wigmore's charts indicate inferences between corresponding claims, where the generalisations used in performing these inferences are not explicitly recorded in the chart. Instead, in their analysis of the case some of the used generalisations are indicated in the text (see e.g. [22, pp. 97–98]). For instance, generalisations used in the inferences from the provided testimonies are of the general form *'If a person testifying under*

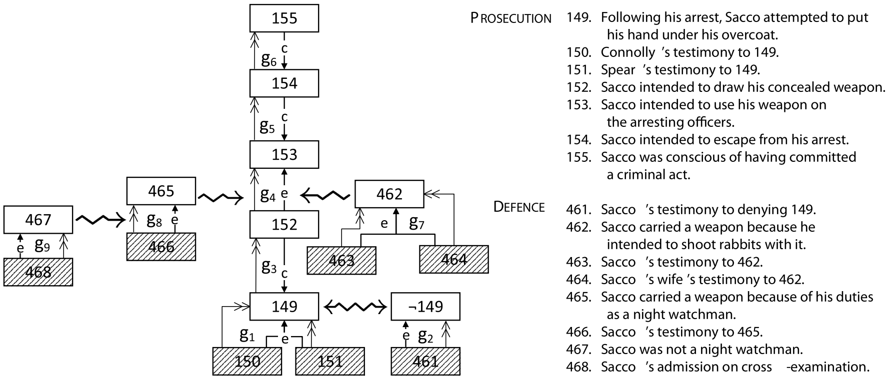


Fig. 19. An IG corresponding to a possible interpretation of the Wigmore chart of Fig. 18, along with the corresponding key list.

oath tells us that event E occurred, then this event (probably, usually, often, etc,) did occur.' [22, p. 88]. As noted by Kadane and Schum [22, pp. 74–76], in constructing their charts abduction is in some instances performed to generate interim hypotheses between the evidence and the ultimate claim Π_3. However, Kadane and Schum do not explicitly indicate which inferences in their charts are abductive and which are deductive.

In their version of Wigmore charts, Kadane and Schum make a distinction between directly relevant and ancillary claims,[4] where the role of an ancillary claim is to show why a generalisation holds or fails in a particular situation [22, p. 53]. Directly relevant and ancillary claims provided by the defence are represented as diamonds and triangles, respectively; for the prosecution, these are represented as circles and squares, respectively. Note that in the Wigmore chart of Fig. 18, all claims provided by the prosecution are directly relevant. All nodes in Kadane and Schum's charts indicate either directly relevant or ancillary claims and nodes corresponding to the evidence are shaded. An arc directed from a node corresponding to an ancillary claim to an arc between two or more claims indicates that this ancillary claim either supports or weakens the applicability of the generalisation in the inference at hand [22, p. 87]. Finally, horizontal lines in the Wigmore chart indicate that information needs to be combined to draw a conclusion.

8.2. Formalising the Wigmore chart as an IG

We now provide a possible formalisation of Kadane and Schum's Wigmore chart of Fig. 18 as an IG. In Fig. 19, an IG is depicted for a possible interpretation of this Wigmore chart. For every claim p in the Wigmore chart, a proposition node p is included in $\mathbf{P}$. In establishing which generalisations could have been used in performing the inferences indicated in the chart, we take the following general approach. In case generalisations are explicitly indicated by Kadane and Schum in the text, then these generalisations are used; otherwise, we first establish whether or not there is a causal relation between the nodes in the chart, and if so, what the direction of causality is. To aid in this process, we determine whether sequences of described events can be interpreted as instances of so-called story schemes [2], which capture stereotypical patterns of causal reasoning. In case p usually/normally/typically causes q, then we establish whether p can be considered the usual cause for q. If this is the case, then evidential generalisation $q \to p$ is included in $\mathbf{G}^{\mathrm{e}}$ to explicitly capture in the IG that p is considered the usual cause of q; otherwise, causal generalisation $p \to q$ is included in $\mathbf{G}^{\mathrm{c}}$ (see also Sect. 2.3).

As noted by Kadane and Schum [22, p. 88], the generalisations used in the inferences from the provided testimonies are evidential (see Sect. 8.1). As propositions 150, 151, 463, 464, 466 and 468 denote testimonies, the IG includes generalisation arcs $g_1: \{150, 151\} \to 149$, $g_7: \{463, 464\} \to 462$, $g_8: 466 \to 465$ and $g_9: 468 \to 467$ in $\mathbf{G}^{\mathrm{e}}$. Here, testimonies 150, 151 and 463, 464 are combined in the antecedents of generalisations g_1 and g_7, respectively, as these sets of propositions concern testimonies to the same claim.

The manner in which claims and links conflict is not precisely specified in Kadane and Schum's Wigmore charts. As we wish to formalise the Wigmore chart of Fig. 18 as an IG, we consider how possible conflicts between claims proposed by the prosecution and defence can be interpreted in terms of the conflict relations defined in Sect. 4.1. As 461 concerns Sacco's testimony to denying 149, proposition $\neg 149$ is included in $\mathbf{P}$, generalisation arc $g_2: 461 \to \neg 149$ is included in $\mathbf{G}^{\mathrm{e}}$, and negation arc $n_1: 149 \leftrightsquigarrow \neg 149$ is included in $\mathbf{N}$.

Kadane and Schum do not indicate which (types of) generalisations were used in performing the inferences between propositions 149 and 155. We note that the inferences between 149 and 155 fit a so-called episode scheme for intentional actions [2, p. 64], a story scheme in which someone's psychological state causes them to form certain goals, which in turn lead to actions that have consequences. In this case, Sacco intended to escape from his arrest (154; goal) as he was conscious of having committed a criminal act (155; psychological state); therefore, we consider 155 to typically cause 154.

[4] Kadane and Schum [22] use the terms 'directly relevant evidence' and 'ancillary evidence'. To avoid confusion with the manner in which the term 'evidence' is used in this paper (i.e. that what has been established with certainty), we instead use the term 'claim'.

Sacco's intention to use his weapon (153) can then be considered a sub-goal of 154 and his intention to draw his concealed weapon (152) a further sub-goal of 153. Sacco's intention to draw his weapon (152) caused Sacco to attempt to put his hand under his overcoat (149; action); more specifically, we consider 152 to typically cause 149. Finally, we consider 153 to be the usual cause for 152, as the usual cause for wanting to draw a weapon is wanting to use this weapon; we therefore include $g_4\colon 152 \to 153$ in $\mathbf{G^e}$. Generalisation arcs $g_3\colon 152 \to 149$, $g_5\colon 154 \to 153$ and $g_6\colon 155 \to 154$ are then included in $\mathbf{G^c}$, as we do not consider their antecedents to express the usual cause for their consequents. Alternatively, it may be argued that some (or all) of these relations are evidential. Below, we show that similar inferences can be performed with the constructed IG and that the same BN graph is constructed from the IG regardless of whether these relations are interpreted as causal or evidential.

In Kadane and Schum's Wigmore chart, it is indicated that 467 is an ancillary claim that weakens (or supports) the applicability of generalisation $g_8\colon 466 \to 465$ in the inference from 466 to 465. In this particular instance, 467 can be interpreted as an exception to generalisation g_8, as the claim that Sacco was not a night watchman indicates that Sacco's veracity in providing his testimony about the reason for carrying a weapon is questionable. Therefore, we include $exc_1\colon 467 \rightsquigarrow g_8$ in **Exc**.

Finally, the conflicts between the defence's claims 462 and 465 and the prosecution's claims 152 and 153 are considered. A possible interpretation is that 462 and 465 indicate exceptions to generalisation $g_4\colon 152 \to 153$ in $\mathbf{G^e}$. Specifically, 462 and 465 can be considered competing alternative explanations for 152: as Sacco carried his weapon for an innocent reason (462 or 465), this caused him to draw his weapon (152) with the intention of surrendering it. In Fig. 19, these exceptions are indicated by curved hyperarcs $exc_2\colon 462 \rightsquigarrow g_4$ and $exc_3\colon 465 \rightsquigarrow g_4$ in **Exc**.

In the Wigmore chart of Fig. 18, the evidence consists of the testimonies; hence, $\mathbf{E_p} = \{150, 151, 461, 463, 464, 466, 468\}$. Given $\mathbf{E_p}$, the inferences that can be read from the IG of Fig. 19 coincide with the inferences indicated in the Wigmore chart. Specifically, given $\mathbf{E_p}$, propositions 149, ¬149, 462, 465 and 467 are deductively inferred from 150 and 151, 461, 463 and 464, 466, and 468 using generalisations g_1, g_2, g_7, g_8 and g_9, respectively. Proposition 152 is then abductively inferred from 149 using g_3, as 149 is deductively inferred. Propositions 153, 154 and 155 are then iteratively inferred using generalisations g_4, g_5 and g_6, respectively.

As mentioned earlier, instead of including causal generalisations $g_3\colon 152 \to 149$, $g_5\colon 154 \to 153$ and $g_6\colon 155 \to 154$, an alternative interpretation is that the antecedents of these generalisations express the usual cause for their consequents; accordingly, evidential generalisations $g'_3\colon 149 \to 152$, $g'_5\colon 153 \to 154$ and $g'_6\colon 154 \to 155$ may instead be included. Similar inferences can then be performed with the constructed IG given $\mathbf{E_p}$; specifically, propositions 152, 153, 154 and 155 are then iteratively deductively inferred given $\mathbf{E_p}$ using g'_3, g_4, g'_5 and g'_6 instead of that some of these inferences are abductive.

8.3. Constructing a BN graph from the IG

We now apply our BN graph construction approach to the IG of Fig. 19 and compare the obtained graph to that of Kadane and Schum.

8.3.1. Applying the BN graph construction approach

By applying our BN graph construction approach to the IG of Fig. 19, the BN graph depicted in Fig. 20b is obtained. By step 1 of our approach, every proposition and its negation are captured as two mutually exclusive values of the same node. Arcs in the BN graph corresponding to generalisation arcs in $\mathbf{G^e} \cup \mathbf{G^c}$ are then directed according to step 2. Additional arcs are then added to $\mathbf{A}_\mathcal{B}$ for every exception arc in **Exc** by step 3 of our approach. Specifically, $exc_1\colon 467 \rightsquigarrow g_8$, $exc_2\colon 462 \rightsquigarrow g_4$ and $exc_3\colon 465 \rightsquigarrow g_4$ are specified in the IG, where $g_8, g_4 \in \mathbf{G^e}$; therefore, additional arcs $467 \to 466$, $465 \to 152$ and $462 \to 152$ are included in $\mathbf{A}_\mathcal{B}$ by step 3a.

Note that in case causal generalisations g_3, g_5 and/or g_6 are replaced by evidential generalisations g'_3, g'_5 and/or g'_6, the same BN graph is obtained by our approach. More specifically, by step 2b arc $\mathrm{Var}(p) \to \mathrm{Var}(q)$ is included for every causal generalisation $g\colon p \to q$, where the same arc is included in $\mathbf{A}_\mathcal{B}$ by step 2a of our approach for every evidential generalisation $g\colon q \to p$.

8.3.2. Comparison to Kadane and Schum's BN graph

The structure of the obtained graph is largely identical to that of the BN graph that Kadane and Schum manually constructed for this part of the case, depicted in Fig. 20c; the differences and similarities between the two BN graphs are now discussed. First, note that Kadane and Schum aggregate nodes 463 and 464 into a single Boolean node K. Similarly, nodes 466, 467 and 468 are aggregated into Boolean node J; possible intercausal effects between 467 and 465 can, therefore, not be explicitly captured in their BN. While aggregation as performed by Kadane and Schum reduces the number of conditional probabilities to be assessed, we prefer to explicitly capture all elements of the IG in the corresponding BN graph to prevent loss of information. The only case in which IG elements are aggregated by our approach is when two propositions p and $\neg p$ appear in the graph, which are then captured as two values of the same node. We note that, by step 6a of our approach, constraints on the CPTs of the BN under construction are automatically obtained, which partially simplifies subsequent probability assessment. Specifically, a head-to-head node is formed in 466, which allows for directly capturing possible interactions between 465 and 467. By step 6a, constraint $\mathbf{X}^-(\{465, 467\}, 466 = \mathit{true})$ is derived on the CPT for node 466. For instance, entries for this CPT can be chosen as follows: $\Pr(466 \mid 465, 467) = 0$,

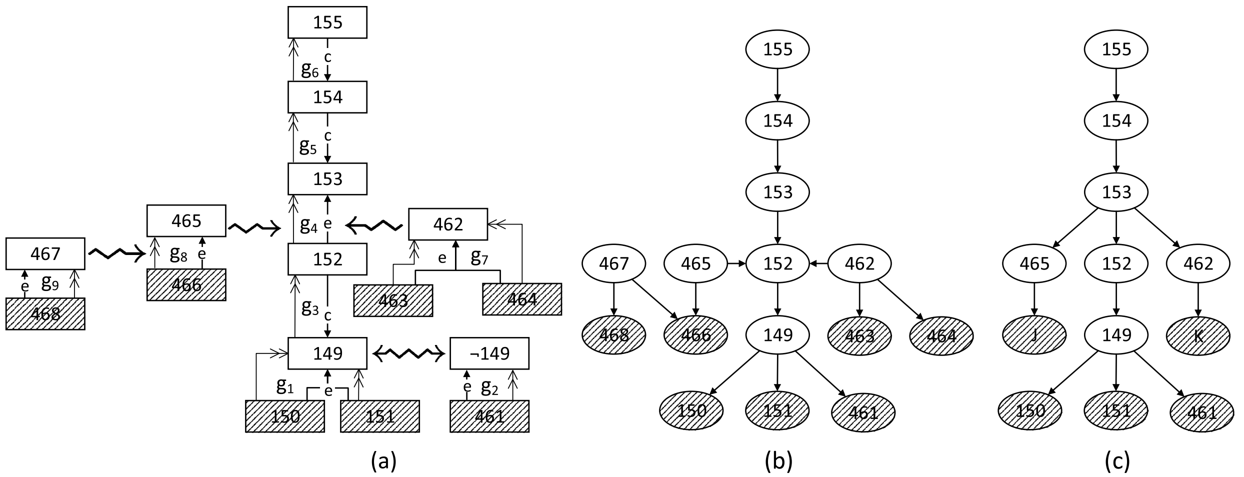


Fig. 20. The IG of Fig. 19 (a); the corresponding BN graph constructed according to our approach (b); adaptation of the BN graph constructed by Kadane and Schum [22, p. 232] (c).

$\Pr(466 \mid \neg 465, \neg 467) = 0.4$, $\Pr(466 \mid 465, \neg 467) = 0.9$, $\Pr(466 \mid \neg 465, 467) = 0.2$, as in this case $0 \cdot 0.4 \leq 0.9 \cdot 0.2$. Note that the conditioned event of conditional probability $\Pr(466 \mid 465, 467)$ cannot actually occur in practice, as Sacco cannot both be and not be a night watchman at the same time. Hence, the exact number to which this conditional probability is set is irrelevant: we choose to set $\Pr(466 \mid 465, 467) = 0$. In case Sacco was indeed a night watchman (467 is not true) but Sacco did not carry a weapon because of this reason (465 is not true), then we find it plausible that Sacco was lying under oath in providing his testimony ($\Pr(466 \mid \neg 465, \neg 467) = 0.4$); more specifically, as he was indeed a night watchman, he can use this as an excuse to claim that he carried his weapon because of this reason. In case Sacco was a night watchman (467 is not true) and Sacco actually carried a weapon because of his duties as a night watchman (465 is true), then we consider the event that Sacco testifies to this claim (466) to be very likely ($\Pr(466 \mid 465, \neg 467) = 0.9$). Finally, in case Sacco was not a night watchman (467 is true) and Sacco did not carry his weapon because of his duties as a night watchman (465 is not true), then we set $\Pr(466 \mid \neg 465, 467) = 0.2$ to again take into account the probability that Sacco may be lying under oath. We believe this probability to be lower than $\Pr(466 \mid \neg 465, \neg 467)$, as we consider it less likely for Sacco to come up with the explanation that he carried his weapon because of his duties as a night watchman if he was in fact not a night watchman.

In the BN graph of Fig. 20b, a head-to-head node is also formed in node 152, which allows for directly capturing possible interactions between 462, 465 and 153. These interactions cannot be captured in the BN graph of Fig. 20c, as in this graph arcs $153 \to 465$ and $153 \to 462$ are included instead of arcs $465 \to 152$ and $462 \to 152$. By step 6*a*, constraints $\mathbf{X}^{-}(\{462, 153\}, 152 = \mathit{true})$, $\mathbf{X}^{-}(\{465, 153\}, 152 = \mathit{true})$ and $\mathbf{X}^{-}(\{465, 462\}, 152 = \mathit{true})$ are derived on the CPT for node 152 in our BN graph. Note that in the BN graph of Kadane and Schum, variables 462 and 465 are conditionally independent from 152 given 153; therefore, in contrast with our BN under construction, for Kadane and Schum's BN it needs to hold that $\Pr(152 \mid 462, 465, 153) = \Pr(152 \mid 153)$. As the entries for the CPT for node 152 in our BN cannot be compared to that of Kadane and Schum, the assessment of the involved probabilities is not further discussed.

We note that for every active chain that exists between two nodes in the BN graph of Fig. 20b given the evidence, there exists an active chain between these nodes in the BN graph of Fig. 20c given the evidence and vice versa; therefore, given $\mathbf{E_V}$, similar probabilistic inferences can be performed in both BN graphs, besides the aforementioned differences. More specifically, as 152 has an instantiated descendant in the BN graph of Fig. 20b, chains between 465 and 462 are active.

Finally, note that the BN constructed from the IG of Fig. 20a cannot be directly used for probabilistic inference. More specifically, the BN is partially specified as only qualitative probabilistic constraints and no exact probabilities are derived on the BN under construction. Moreover, the derived qualitative probabilistic constraints are only a subset of those required for the specification of a QPN [33] (see also Sect. 9.3). The derived qualitative probabilistic constraints may serve as input for a subsequent elicitation procedure for obtaining a fully specified QPN or BN for (qualitative) probabilistic inference.

8.4. Concluding remarks

In this section, we have performed a first validation of our BN graph construction approach by means of a case study. We have provided a possible interpretation of Kadane and Schum's Wigmore chart as an IG, which illustrates that the IG-formalism is sufficiently expressive to model a complex case in a precise way. We have then applied our approach to the constructed IG. Upon comparing the BN graph obtained by applying our approach to the BN graph that Kadane and Schum manually constructed, we have concluded that the graphs are largely identical and that similar probabilistic inferences can be performed for the case at hand. As Kadane and Schum provided a thorough and extensive probabilistic analysis of the case, these similarities are a positive result of our validation and offer a first indication that BNs constructed from IGs by our approach are of good quality. Moreover, the differences obtained illustrate that our approach may provide a more principled way of constructing BN graphs than the manner in which Kadane and Schum constructed their BNs. In particular, Kadane and Schum in some cases aggregated multiple claims in the Wigmore chart into single nodes in the BN graph, while by applying our approach all elements of the IG are explicitly captured in the corresponding BN graph to prevent loss of information. Furthermore, in comparison to the BN graph of Kadane and Schum head-to-head nodes are formed in our BN graph, which allows for directly capturing possible interactions between nodes in the graph.

9. Related research

In this section, related research on inference with causality information (Sect. 9.1), BN graph construction (Sect. 9.2), the relations between probability and inference (Sect. 9.3), and intermediary formalisms (Sect. 9.4) are discussed.

9.1. Inference with causality information

In this paper, we have presented the graph-based IG-formalism for deductive and abductive inference with causal and evidential information. As mentioned earlier, the currently presented IG-formalism is a further specification of the IG-formalism that appeared in our previous work [42]. More specifically, in the current paper we define a number of new concepts, namely negation arcs (Definition 6), competing alternative explanations (Definition 11) and (restrictions on) generalisation chains (Definitions 4 and 5 and p. 255), and we now explicitly describe a number of concepts in the form of definitions and remarks, including prediction (Remark 1), ambiguous inference (Remark 2) and alternative causes (Definition 3). Our IG-formalism extends on a preliminary formalism used in [39,40], in which the roles of generalisation and inference are not separated; therefore, this preliminary formalism does not provide a precise enough account of reasoning with causal and evidential information.

Most related formalisms for inference with causal and evidential information are logic-based instead of graph-based [2,4,21,29,35]. Poole's Theorist framework [29] allows for both deductive and abductive inference, which is established using only causal defaults. Complications with reasoning using both causal and evidential defaults as identified by Pearl [27] are thus avoided. In the hybrid theory proposed by Bex [2], deductive and abductive inference are used in constructing evidential arguments and causal stories. Compared to our IG-formalism, the hybrid theory does not allow for most types of mixed inference and largely avoids the problems associated with mixed inference as identified by Pearl [27]. Building on his hybrid theory, Bex proposed his integrated theory of causal and evidential arguments [4]. In Bex' integrated theory, the roles of generalisation and inference are not separated; instead, causal and evidential inference rules are defined and arguments are constructed by forward chaining such inference rules. Actual abductive inference is thus not performed by constructing arguments.

As noted in Sect. 7.2, inference chains in IGs are comparable to arguments as defined in ASPIC$^+$ [30]. Besides the mentioned distinctions between these formalisms, our graph-based IG-formalism deviates from the logic-based ASPIC$^+$ framework as we introduce a new type of conflict, namely conflict between competing alternative explanations, and impose constraints on the different types of inferences that may be performed with the different types of generalisations.

Graph-based formalisms for reasoning with causality information have also been proposed, notably Pearl's causal diagrams [28]. Pearl provides a framework for causal inference in which diagrams are queried to determine if the assumptions available are sufficient for identifying causal effects. Compared to our IG-formalism, the aim of this framework is different in that it serves to identify causality instead of providing a way to reason with modelled causal knowledge. Furthermore, causal diagrams require probabilistic quantification to be queried, while IGs are qualitative.

9.2. BN graph construction

To facilitate BN graph construction, construction methods have been proposed in the literature. As noted in the introduction, work on the construction of BNs from information represented in ontologies [17,31] is related to our BN construction approach based on IGs. In the approach of Fenz [17], an initial BN graph is automatically constructed after a manual selection of relevant concepts and relations from an OWL ontology. Specifically, concepts are mapped to nodes in the BN graph and the direction of the relation between two concepts is used to direct arcs between corresponding nodes in the graph as a first heuristic. However, properties regarding the represented independence relation are not investigated; instead, Fenz notes that the obtained BN graph needs to be verified and refined manually by the BN modeller. Ramírez-Noriega and colleagues [31] proposed a similar approach in the domain of intelligent tutoring systems, where the focus lies on obtaining the quantitative part of the BN.

In other related work, Bex and Renooij [5] identified constraints on BNs given arguments, based on the inferences on which arguments are built and the existing conflicts between arguments. These constraints suffice for constructing an undirected skeleton of a BN graph. In their approach, ASPIC$^+$ [30] is taken as a starting point for BN graph construction; for reasons mentioned in Sect. 9.1, we wish to refrain from using ASPIC$^+$ as an intermediary formalism. In our previous work, we explored the possibility of BN construction from a graph-based intermediary formalism using a preliminary version of IGs in [39,40]. As mentioned in Sect. 9.1, this preliminary IG-formalism does not provide a precise enough account of reasoning with causal and evidential information; hence, the BN construction approaches provided in these papers are also imprecise.

Throughout the literature, many (often domain-specific) BN fragments and modules, also called idioms, have been proposed. In the legal domain, Fenton and colleagues [16, Ch. 13] proposed BN fragments to model recurring patterns of legal reasoning, such as structures for corroboration. Laskey and Mahoney [25] proposed BN fragments in the domain of military situation assessment, and studied how fragments can be combined to construct more complex networks. In contrast with these manual fragment-based approaches for BN graph construction, our approach allows for automatically constructing an initial BN graph from an IG that satisfies a number of desirable properties, for instance regarding the represented independence relation, where generalisations and conflicts can be incorporated and combined in an IG without having to conform to any predefined pattern or configuration. Arguably this allows our BN construction approach to be applied more flexibly in practice, a claim that should be empirically evaluated in future work.

To facilitate incremental BN construction, Object-Oriented BNs (OOBNs) were introduced by Koller and Pfeffer [24]. With OOBNs, it becomes possible to incrementally construct a BN top-down, using fragments and modules such as proposed throughout the literature to gradually construct a network. Unlike our approach, OOBNs do not provide an automated way of constructing BN graphs; instead, OOBNs allow experts to more quickly construct a BN manually by allowing recurrent patterns to be incorporated. The concept of reusable network fragments was also the basis of Hypothesis Management Frameworks (HMFs) proposed by Van Gosliga and van de Voorde [19], which are generally applicable and not intended for a specific domain. With HMFs, a modular approach is taken, enabling the specification of details about a case without losing perspective on the case as a whole.

9.3. The relations between probability and inference

As discussed by Bex and Renooij [5], the exact probabilistic interpretation of inference and evidence, and hence the various types of constraints on a BN, is a contentious issue. The interpretation of strict inferences is straightforward: the consequent is necessarily true given the antecedents. However, with respect to defeasible inferences different ideas exist on how they should be modelled probabilistically. For instance, for a defeasible inference from $p_1, \ldots, p_n$ to q it can be assumed that $\Pr(q \mid p_1, \ldots, p_n) > 0$; however, this constraint is rather weak. Another possible interpretation is that this probability should be greater than 0.5 in that the antecedents make it more likely than not that the consequent is true. Yet another reading is that the probability of the consequent given the antecedents should be greater than the prior probability of the consequent, which is a Bayesian interpretation explored by for instance Crupi and colleagues [10]. This interpretation is based around the notion of conditional independence: in case $\Pr(q \mid p_1, \ldots, p_n) = \Pr(q)$, q is conditionally independent from $\{p_1, \ldots, p_n\}$, and learning that $\{p_1, \ldots, p_n\}$ is the case will be uninformative to q. Therefore, according to this interpretation it should hold that $\Pr(q \mid p_1, \ldots, p_n) > \Pr(q)$.

Another possible probabilistic interpretation is to view inferences that can be read from an IG given the evidence as qualitative influences [13]. Specifically, variable P is said to have a positive qualitative influence on variable Q if $\Pr(q \mid p) \geq \Pr(q \mid \neg p)$ and a negative qualitative influence if $\Pr(q \mid p) \leq \Pr(q \mid \neg p)$. Interpreting all inferences between propositions $p_1, \ldots, p_n$ and q that can be read from an IG as positive qualitative influences and all inferences between propositions $p_1, \ldots, p_n$ and $\neg q$ as negative qualitative influences, a fully specified qualitative probabilistic network (QPN) may be constructed by our approach which can be used for qualitative probabilistic inference [33]. Quantification of QPNs can then be performed incrementally by specifying probability intervals for CPTs for nodes in the graph as an intermediary step, resulting in so-called semi-qualitative probabilistic networks [33] that can also be used for probabilistic inference. Alternatively, a credal network [9] can be constructed [7].

The point of the above discussion is that there are many probabilistic interpretations of inference and evidence, and choosing exactly which ones to use is not trivial. One way to deal with discussions involving probabilistic constraints is to allow one to reason *about* such constraints [23,41]. Specifically, the approaches proposed in [23,41] allow domain experts to reason and argue about BN modelling decisions, where computational argumentation [15] is used to resolve disagreements as much as possible.

9.4. Intermediary formalisms

Our IG-formalism serves as an intermediary formalism between analyses performed using informal reasoning tools and formal AI systems such as argumentation systems (see [42]) and BNs. Viewed this way, in the context of argumentation the IG-formalism is comparable to the Argument Interchange Format (AIF) [3], an argumentation ontology that serves as an intermediary formalism between analyses performed using argument diagramming tools [1,26] and formal argumentation frameworks such as the ASPIC$^+$ framework [30].

In the context of BNs, another graph-based intermediary formalism was proposed by Timmer and colleagues [37]. They introduced the support graph that captures general reasoning patterns represented by a BN for the purpose of explaining such patterns in terms of argumentation. Compared to the present paper, Timmer and colleagues' work is in the reverse direction, namely from BNs to domain-specific rules and inferences (i.e. arguments).

10. Conclusion and future research

In this paper, we have presented the IG-formalism, which provides a precise account of the interplay between deductive and abductive inference and causal and evidential generalisations and which imposes constraints on the inferences that may be performed with this knowledge given the evidence. Moreover, we have introduced a BN graph construction approach that demonstrates that the knowledge expressed in an IG, namely the generalisations and conflicts expressed in the graph, can be directly exploited for this purpose. Given the evidence, sequences of inferences can be read from an IG; we have formally proven that all such sequences in an IG are captured in the form of induced active chains in the corresponding BN graph constructed by our approach, as intended. Moreover, by considering the inferences that can be read from an IG given the evidence, some qualitative constraints on the (conditional) probabilities of the BN under construction are derived. These qualitative probabilistic constraints may serve as input for a subsequent elicitation procedure for obtaining a fully specified QPN [33] or BN for (qualitative) probabilistic inference. We have identified conditions on the IG under which the fully automatically constructed initial graph is guaranteed to be a DAG, simplifying the (manual) verification step. Lastly, we have identified bounds on the size of the CPTs and the complexity of probabilistic inference in BNs constructed by our approach.

Our IG-formalism, together with our BN construction approach, allow us to construct an initial BN graph from a domain expert's initial analysis, capturing similar reasoning patterns as can be read from their IG given the evidence; it thereby simplifies the BN elicitation process. We note that BN construction is an iterative process in which both the domain expert and BN modeller should stay involved; this also holds when applying our approach, as even the provided IG may be incomplete or may be subject to change over time. To aid in this iterative process, approaches can be used [23,41] such as discussed in Sect. 9.3.

IGs formalise analyses performed by domain experts using the informal reasoning tools they are familiar with. In interpreting a performed analysis as an IG, an additional knowledge elicitation step in consultation with the domain expert may be required as the used generalisations and the manner of conflict are typically left implicit in these analyses. IGs may also be directly constructed by domain experts in case work. As mentioned earlier, we expect direct IG construction to be more straightforward than direct BN construction for domain experts unfamiliar with the BN-formalism, a claim we intend to empirically evaluate in our future work.

In future work general guidelines for IG construction may be formulated. In our case study of Sect. 8 we constructed an IG corresponding to a Wigmore chart according to a number of general heuristics. For instance, in establishing which generalisations could have been used in constructing the chart we among other things determined whether sequences of described events could be interpreted as instances of story schemes [2] (see Sect. 8.2). In future work a database of schemes that capture general patterns of defeasible reasoning (including argumentation schemes [1] and story schemes) may be composed, instantiations of which can be used as building blocks in facilitating IG construction. Such an approach would in turn facilitate BN graph construction. In the context of BNs such an approach is comparable to the idiom-based approaches to BN construction discussed in Sect. 9.2. We also intend to increase the expressivity of our IG-formalism by allowing generalisations that are neither causal nor evidential. For instance, definitions, or abstractions [8] allow for reasoning at different levels of abstraction, such as stating that guns can generally be considered deadly weapons. Another example of a different type of generalisation is a generalisation representing a mere statistical correlation, such as a correlation between homelessness and criminality. In the manual construction of BN graphs, arcs are typically directed using the notion of causality as a guiding principle; however, non-causal relations are also considered in the literature. For instance, in the BN construction guidelines of Fenton and colleagues [16, Ch. 7] not only causal but also definitional relations are considered, in which arcs in the BN graph are oriented in the direction in which a sub-attribute (or combination of sub-attributes) defines an attribute. In previous research, we investigated BN graph construction from a preliminary form of IGs including abstractions and other types of generalisations [40]; in our future work, we intend to generalise the currently proposed BN graph construction approach to an extended IG-formalism allowing for such generalisations.

We will also focus on deriving more probabilistic constraints such as qualitative influences corresponding to inferences that can be read from an IG given the evidence. Our approach may then serve to construct fully specified QPNs for qualitative probabilistic inference [33], as discussed in Sect. 9.3. We also intend to evaluate our approach by assessing the quality of BNs constructed from IGs. Since we are considering BN construction in data-poor domains, we assume that there is insufficient data to learn a reliable BN from and that such a BN is therefore not available for comparison. A quality assessment should therefore mainly be based on compliance with best practice guidelines for BN construction [16, Ch. 7].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F. Bex, H. Prakken, C.A. Reed, D. Walton, Towards a formal account of reasoning about evidence: argumentation schemes and generalisations, Artif. Intell. Law 11 (2–3) (2003) 125–165.
[2] F. Bex, Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory, Springer, 2011.
[3] F. Bex, S. Modgil, H. Prakken, C.A. Reed, On logical specifications of the argument interchange format, J. Log. Comput. 23 (5) (2013) 951–989.
[4] F. Bex, An integrated theory of causal stories and evidential arguments, in: Proceedings of the Fifteenth International Conference on Artificial Intelligence and Law, ACM Press, 2015, pp. 13–22.
[5] F. Bex, S. Renooij, From arguments to constraints on a Bayesian network, in: P. Baroni, T.F. Gordon, T. Scheffler, M. Stede (Eds.), Computational Models of Argument: Proceedings of COMMA 2016, IOS Press, 2016, pp. 95–106.

[6] S.W. van den Braak, H. van Oostendorp, H. Prakken, G.A.W. Vreeswijk, Representing narrative and testimonial knowledge in sense-making software for crime analysis, in: E. Francesconi, G. Sartor, D. Tiscornia (Eds.), Legal Knowledge and Information Systems: JURIX 2008: The Twenty-First Annual Conference, IOS Press, 2008, pp. 160–169.
[7] C.P. de Campos, F. Cozman, Belief updating and learning in semi-qualitative probabilistic networks, in: F. Bacchus, T. Jaakkola (Eds.), Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence, AUAI Press, 2005, pp. 153–160.
[8] L. Console, D.T. Dupré, Abductive reasoning with abstraction axioms, in: G. Lakemeyer, B. Nebel (Eds.), Foundations of Knowledge Representation and Reasoning, vol. 810, Springer, 1994, pp. 98–112.
[9] F.G. Cozman, Credal networks, Artif. Intell. 120 (2000) 199–233.
[10] V. Crupi, K. Tentori, M. Gonzalez, On Bayesian measures of evidential support: theoretical and empirical issues, Philos. Sci. 74 (2) (2007) 229–252.
[11] A. Darwiche, Modeling and Reasoning with Bayesian Networks, Cambridge University Press, 2009.
[12] A.P. Dawid, Beware of the DAG!, in: I. Guyon, D. Janzing, B. Schölkopf (Eds.), NIPS Causality: Objectives and Assessment, 2010, pp. 59–86.
[13] M.J. Druzdzel, M. Henrion, Intercausal reasoning with uninstantiated ancestor nodes, in: D. Heckerman, E. Mamdani (Eds.), Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, 1993, pp. 317–325.
[14] M.J. Druzdzel, L.C. van der Gaag, Elicitation of probabilities for belief networks: combining qualitative and quantitative information, in: P. Besnard, S. Hanks (Eds.), Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, 1995, pp. 141–148.
[15] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artif. Intell. 77 (2) (1995) 321–357.
[16] N. Fenton, M. Neil, Risk Assessment and Decision Analysis with Bayesian Networks, CRC Press, 2012.
[17] S. Fenz, An ontology-based approach for constructing Bayesian networks, Data Knowl. Eng. 73 (2012) 73–88.
[18] L.C. van der Gaag, E.M. Helsper, Experiences with modelling issues in building probabilistic networks, in: A. Gómez-Pérez, V.R. Benjamins (Eds.), International Conference on Knowledge Engineering and Knowledge Management, vol. 2473, Springer, 2002, pp. 21–26.
[19] S.P. van Gosliga, I. van de Voorde, Hypothesis management framework: a flexible design pattern for belief networks in decision support systems, in: S. Renooij, H.J.M. Tabachneck-Schijf, S.M. Mahoney (Eds.), Proceedings of the Sixth Bayesian Modelling Applications Workshop at the Conference on Uncertainty in Artificial Intelligence, CEUR, 2008.
[20] F.V. Jensen, T.D. Nielsen, Bayesian Networks and Decision Graphs, 2nd ed., Springer, 2007.
[21] J.R. Josephson, S.G. Josephson, Abductive Inference: Computation, Philosophy, Technology, Cambridge University Press, 1994.
[22] J.B. Kadane, D.A. Schum, A Probabilistic Analysis of the Sacco and Vanzetti Evidence, John Wiley & Sons Inc., 1996.
[23] J. Keppens, On modelling non-probabilistic uncertainty in the likelihood ratio approach to evidential reasoning, Artif. Intell. Law 22 (3) (2014) 239–290.
[24] D. Koller, A. Pfeffer, Object-oriented Bayesian networks, in: D. Geiger, P. Shenoy (Eds.), Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, 1997, pp. 302–313.
[25] K.B. Laskey, S.M. Mahoney, Network fragments: representing knowledge for constructing probabilistic models, in: D. Geiger, P.P. Shenoy (Eds.), Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, 1997, pp. 334–341.
[26] A. Okada, S.J. Buckingham Shum, T. Sherborne (Eds.), Knowledge Cartography: Software Tools and Mapping Techniques, 2nd ed., Springer, 2014.
[27] J. Pearl, Embracing causality in default reasoning, Artif. Intell. 35 (2) (1988) 259–271.
[28] J. Pearl, Causality: Models, Reasoning, and Inference, 2nd ed., Cambridge University Press, 2009.
[29] D. Poole, Representing diagnosis knowledge, Ann. Math. Artif. Intell. 11 (1–4) (1994) 33–50.
[30] H. Prakken, An abstract framework for argumentation with structured arguments, Argum. Comput. 1 (2) (2010) 93–124.
[31] A. Ramírez-Noriega, R. Juárez-Ramírez, J.J. Tapia, V.H. Castillo, S. Jiménez, Construction of conditional probability tables of Bayesian networks using ontologies and Wikipedia, Comput. Sist. 23 (4) (2019) 1275–1289, https://doi.org/10.13053/CyS-23-4-2705.
[32] R. Reiter, A logic for default reasoning, Artif. Intell. 13 (1–2) (1980) 81–132.
[33] S. Renooij, L.C. van der Gaag, From qualitative to quantitative probabilistic networks, in: A. Darwiche, N. Friedman (Eds.), Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence, UAI, Morgan Kaufmann, 2002, pp. 422–429.
[34] A. de Ronde, B. Kokshoorn, C.J. de Poot, M. de Puit, The evaluation of fingermarks given activity level propositions, Forensic Sci. Int. 302 (2019) 109904.
[35] M. Shanahan, Prediction is deduction but explanation is abduction, in: N.S. Sridharan (Ed.), Proceedings of International Joint Conference on Artificial Intelligence 89, Morgan Kaufmann, 1989, pp. 1055–1060.
[36] F. Taroni, C.G. Aitken, P. Garbolino, A. Biedermann, Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science, Wiley, 2014.
[37] S.T. Timmer, J.-J.C. Meyer, H. Prakken, S. Renooij, B. Verheij, A two-phase method for extracting explanatory arguments from Bayesian networks, Int. J. Approx. Reason. 80 (2017) 475–494.
[38] N. Timmers, The hybrid theory in practice: a case study at the Dutch police force, Master's thesis, Utrecht University, The Netherlands, 2017.
[39] R. Wieten, F. Bex, H. Prakken, S. Renooij, Exploiting causality in constructing Bayesian networks from legal arguments, in: M. Palmirani (Ed.), Legal Knowledge and Information Systems. JURIX 2018: The Thirty-First Annual Conference, vol. 313, IOS Press, 2018, pp. 151–160.
[40] R. Wieten, F. Bex, H. Prakken, S. Renooij, Constructing Bayesian network graphs from labeled arguments, in: G. Kern-Isberner, Z. Ognjanović (Eds.), Proceedings of the Fifteenth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU, in: LNAI, vol. 11726, Springer, 2019, pp. 99–110.
[41] R. Wieten, F. Bex, H. Prakken, S. Renooij, Supporting discussions about forensic Bayesian networks using argumentation, in: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, ACM Press, 2019, pp. 143–152.
[42] R. Wieten, F. Bex, H. Prakken, S. Renooij, Deductive and abductive reasoning with causal and evidential information, in: H. Prakken, S. Bistarelli, F. Santini, C. Taticchi (Eds.), Computational Models of Argument: Proceedings of COMMA 2020, IOS Press, 2020, pp. 383–394.
[43] J.H. Wigmore, The Principles of Judicial Proof, Little, Brown and Company, 1913.
[44] J. de Zoete, M. Sjerps, D. Lagnado, N. Fenton, Modelling crime linkage with Bayesian networks, Sci. Justice 55 (3) (2015) 209–217.